EXECUTIVE DECISION MAKING

Third Printing

Contributors

This text is a synthesis of the work of many Executive Decision Making faculty members of the of the National Security Decision Making Department, past and present. The current Executive Decision Making faculty is:

Prof. Thomas R. Fedyszyn
Prof. William C. Martel
LtCol Norman E. Hitchcock, USMC
Prof. H. Marshall Hoyler
Col Charles Murray, USAF
LtCol Alan J. Neff, USMC
Lt Col Cynthia S. Perroti, USAF
Prof. Charles R. Sipe Jr
CDR Raymond E. Sullivan, USN
CDR Andrew N. Wall, USN

Cover: U.S. Naval War College Graphics Department

EXECUTIVE DECISION MAKING

6th Edition
1 February 2002
Colonel Charles H. Murray, U.S. Air Force
Editor

NATIONAL SECURITY DECISION MAKING DEPARTMENT
U.S. NAVAL WAR COLLEGE
NEWPORT, R.I.
This publication is also available on the U.S. Naval War College website:
<http://www.nwc.navy.mil/nsdm/nsdmedm2.htm> (PDF)

TABLE OF CONTENTS

TABLE OF FIGURES

PREFACE

THIS IS THE SECOND TEXTBOOK that supports the Executive Decision Making Course of the National Security Decision Making Department of the U.S. Naval War College. The first text, *Resource Allocation: The Formal Process* describes the people, processes and products that establish our national strategies, determine the nature of our force structure, propose defense programs and policies, and see them through the federal budget process. The formal process is the programmatic and budgetary machinery of the Office of the Secretary of Defense, the services, the Chairman of the Joint Chiefs of Staff, the unified commands, and the defense agency staffs that many of us will someday join.

This text is concerned with the skills defense executives need to solve complex problems and make good decisions. We believe that critical and structured thinking—a combination of high quality analysis, rationality, and professional judgment—form the best approach for solving complex problems. To help us equip our students to think critically, this book is built around a decision-making framework that emphasizes analysis, the coin of the realm in defense resource allocation; indeed chapters 3 through 9 are a field guide to analysis as currently practiced in the Department of Defense.

The Executive Decision-Making Framework in Appendices 1 and 2 is the backdrop for this book because adding structure to problem solving greatly facilitates selecting the best alternative by allowing us to approach the parts of a problem incrementally. The book follows the framework through its phases: Definition, Analysis, Decision, Reconciliation, and Execution. The framework, although typical in many ways of numerous decision-making frameworks, is exceptional because it places a high premium on reconciling a decision with other interested parties. Despite its hierarchical structure, the Department of Defense is very much a consensus-oriented institution. Negotiating skills become increasingly important as officers and civilian defense professionals become more senior; therefore we have devoted a separate chapter and significant course time to reconciliation.

This sixth edition text was originally and primarily authored by CDR Doug Hancher and is the successor to previous efforts by the Executive Decision Making faculty (under a variety of names). It retains the same overarching themes. The first is the emphasis on rational analysis begun by Dr. Warren Rogers in the early 1970s when Admiral Stansfield Turner dramatically reshaped our war college. Later that decade, Dr. Bill Turcotte elevated our analytical approach to the executive level and emphasized its linkage to decision making. The third major theme is the insertion of reconciliation that was pioneered by Dr. Steve Fought at the end of the 1980s. Dr. Ken Watman increased our emphasis on rationality and decision-making theory and added our examination of risk and uncertainty.

Prof. Andy Mackel produces the seminal third edition of this book, single-handedly taking the text into the computer age, moving it past an amalgamation of loosely bound readings. Without their efforts, and those of the rest of the faculty, past and present, this text would not have its current form.

Charles H. Murray
Colonel, USAF

CHAPTER 1

MAKING DEFENSE DECISIONS

Reason and free enquiry are the only effectual agents against error.
—Thomas Jefferson, Notes on the State of Virginia, c. 1781-1783

MAKING HIGH-LEVEL DEFENSE DECISIONS is a large part of being a senior military officer or career defense civilian. Earlier in your professional life, many of your decisions concerned near-term problems involving small numbers of people and a limited array of resources. You could usually make these decisions by using standard procedures or by relying on your personal experience. Now, increasingly, you will find yourself making or participating in more complex decisions that affect the long-term capabilities of your organization and therefore the welfare of the nation—issues that concern force structure, organization, modernization, operations, and policy.

These kinds of choices will push you into new, unfamiliar circumstances in which procedure and experience are no longer sufficient unto themselves. How do you decide whether to advocate producing a next-generation weapon or to push instead for a complete technological leap forward? This text will help you answer that kind of question by providing a structured approach to problem solving and decision making. It will help you identify and bound not only what is known and unknown, but also what ingredients are necessary to make a good decision. Furthermore, we will discuss analysis, a principal tool that senior leaders use to help identify the most rational course of action. Our ultimate aim is to equip you with the tools to make the intelligent choices that are so necessary for building the right forces for the future.

Sometimes making a good decision is easy: there are problems and decisions in which the best alternative is obvious, the proverbial "no-brainer." Defense decisions, however, are almost always far more difficult. Why is this so? First, a decision may require that we simultaneously consider many interrelated factors, forcing us to decide which are critical. Often we must balance choices between modernization, force structure, readiness, and infrastructure in an environment of constrained resources that does not permit us to do everything for national security we think we should. Second, a decision may involve substantial risks, especially when there is uncertainty and the consequences of the decision are not entirely predictable. Indeed, uncertainty and the risk it produces dominates force structure decisions in particular. Finally, defense decisions almost always involve conflicting objectives and perspectives. We make decisions ourselves, personally and within our own organizations, and then we acknowledge that each element of the defense community affected by this decision has its own interests and perspectives, which may conflict with our own. We must identify which objectives we will satisfy, which we will not, and by how much.

How do we make such decisions? Judgment and professional experience—our own and that of others—are of course invaluable in making sound and effective defense decisions. But, for reasons we will examine, intuition, experience, and judgment alone may not suffice, especially when decisions involve new and unfamiliar situations. We require a more rigorous, objective, and systematic approach to augment our experience. We call this approach Analysis-Based Decision Making. It can be taught and learned like most of the other skills crucial to good leadership, and when combined with professional judgment, provides a powerful tool.

Procedure, Experience, and Analysis

The methods we choose for making complex defense decisions will have much to do with the success of our decisions. We therefore ask: Are some methods more effective than others and in which situations are they most beneficial? We will consider each of the three main approaches to defense decision making. Keep in mind these three approaches are not mutually exclusive. The skillful decision maker selects the method best suited to the problem and often blends techniques together.

PROCEDURE-BASED DECISION MAKING

We are all familiar with doing it "by the book." Procedure-based decision making relies upon a body of explicit instructions for guidance in choosing a course of action. The instructions may take the form of standard operating procedures; checklists; tactics, techniques, and procedures; doctrine; manuals; laws and regulations; planning factors; etc. At its heart, procedure-based decision making consists of pattern matching. As we encounter a problem, we compare it with similar problems that we encountered during training. When we find a good match, we apply the solution, detailed in procedures, that we were trained to apply.

Procedure-based decision making has several strengths. First, this approach allows many individuals to benefit from the best knowledge available without having to repeat the mistakes of others who encountered the same problem in the past. At their best, these procedures are the distilled wisdom of intelligent and careful people who have systematically arrived at optimal solutions. Second, this technique introduces predictability and uniformity into the way a large organization deals with standard recurring problems and that, in turn, increases the coherence and focus of the organization. Third, this approach permits us to make complex decisions rapidly when we are severely constrained by time. Finally, by relying on procedures, we do not need to seek specialists for a particular type of problem every time such a problem arises. Generalists trained to use a relevant body of guidance are capable of resolving many complex situations without fully understanding the underlying substance of the problem—just as we operate complex technology daily without understanding much of its inner workings.

Not surprisingly, procedure-based decision making is most effective when we apply it to problems that arise repeatedly in more or less the same form, when the time to decide is short, and when our desire to ensure a uniform response is high. Nuclear engineers and pilots use checklists because avoiding even small omissions or deviations is important. We also rely heavily on procedures in combat. Here time pressures for decisions are extreme and coordination among units requires mutually predictable behavior and responses to problems. The fundamental purpose of combat training and doctrine, at the individual and small-unit levels, is to ingrain individuals with a set of reflexes designed to enable them to recognize different types of combat problems and to react appropriately and predictably under even the most exceptional conditions.[1]

The fundamental weaknesses of procedure-based decision making are its requirement for problems that can be easily categorized and its need for a body of relevant and effective guidance to solve each problem. The decision has to be foreseen or experienced by the creators of the procedures. As long as each decision situation we encounter is close enough to one of those addressed in the procedures, we can effectively rely on them. But what happens when the decision is sufficiently different or complex so that it becomes difficult to know which procedure to use – or whether any of them are applicable? The more senior you become, the more likely it is that the complex problems requiring your attention will be new or unique and fall beyond the ken of procedure-based decision making.

Taking the problem a step further, what happens when an individual in the habit of relying on procedures addresses *all* problems using procedures? As the old saw says, "To a man with a hammer, every problem looks like a nail." We have all encountered these kinds of people, especially when dealing with bureaucracies; such behavior is counterproductive and frustrating. In sum, so long as a decision fits into the space covered by procedures, procedure-based decision making is effective. The more the problem involves nonstandard factors and issues, the less likely our reliance on procedures will be productive.

THE ROYAL NAVY'S FIGHTING INSTRUCTIONS: PROCEDURE BECOMES DOGMA

A set of procedures may outlive their usefulness as the British Royal Navy discovered in the nineteenth century with its archaic *Fighting Instructions*. At the beginning of the seventeenth century, as large groups of sail-powered warships met in battle, their dominant armament was cannon, most of which were mounted to fire athwartship. To clear their own fields of fire, line-ahead battle formations, wherein one ship followed another, became the norm for fleets. These ships were optimized to sail downwind (square-rigged) and therefore they performed very poorly as they sailed with their bows pointed closer into the wind. Fleets maneuvered strenuously to gain an upwind positional advantage over their opponents from which they could sail down to accept combat or remain upwind to decline it: their adversaries could not effectively sail up to reach them. Ideally, starting from the upwind position, admirals would try to "cross the T" of their foes with their line of battle, bringing the broad-side weight of their ships' guns to bear, putting an unanswerable raking fire on the exposed ends of the enemy ships. More often, neither side crossed the T and the lines of battle converged until one side was destroyed or broke and fled. Such simple tactics and maneuvers were a necessity with these older, clumsier sailing ships.

The Royal Navy codified these tactics for their officers in their *Fighting Instructions*. This doctrine spelled out which ships were to be placed where in the line of battle and charged each captain to maintain the line of battle without breaks, regardless of the peril to his own ship. The Admiralty de-

1. The importance of predictable behavior in combat is reflected in a current dilemma within NATO. The U.S. Army has adopted a very violent and rapid maneuver doctrine, AirLand Battle, which employs quick concentrations of mass and fires and then dispersion and maneuver, often disregarding traditional front lines and concern for its flanks. Other NATO ground formations, often filled with short-term conscripts and lacking the command and control technology of U.S. forces, are disconcerted that they will not be able to keep up with the tempo of U.S. operations. They are further concerned that their flanks will be exposed to turning movements without continuous allied front lines.

signed these procedures to choreograph fighting a naval battle to the point where every officer in the Royal Navy knew his role and duty regardless of how long he had been under a particular superior. For decades, then centuries, Royal Navy admirals and captains upon pain of court-martial placed their ships in a line of battle parallel to the enemy's and closed the range for a slugging duel.

The leadership of the Senior Service required slavish obedience to the *Fighting Instructions.* For example, after Matthew's disappointing action against an equivalently sized allied Spanish and French fleet off Toulon in 1774, the admiral, his deputy, and 11 of 29 ships' captains were court-martialed. Matthew was cashiered for breaking the line as he maneuvered toward the enemy fleet. His deputy, who failed to support Matthew's charge with anything other than a distant cannonade, was acquitted by the court-martial because he had doggedly maintained the line of battle.[2] As ship's sails and rigging improved throughout the eighteenth century and ships became more maneuverable, the *Fighting Instructions* remained unchanged.

Accidentally at first, beginning with Rodney at the Battle of the Saints (1782) and then, as improvised by Duncan at Campertown (1797), individual British captains and admirals began winning dramatic battles with tactics contrary to the *Fighting Instructions*. As a ship's captain at the Battle of Cape St. Vincent in 1797, Nelson broke the line of battle and charged the enemy, accompanied impulsively (and fortunately) by several of his colleagues. Their action decided the outcome favorably for the British. As the fleet commander at Trafalgar in 1805, Nelson's battle plan for a perpendicular approach toward the enemy line openly contradicted the procedures in the *Fighting Instructions*. Furthermore, he unconventionally decentralized tactical control of his fleet and ships' maneuvers. After making his overall intentions for the battle clear, Nelson left his captains to their discretion. The French and Spanish lost 17 of their 33 ships of the line to the 27 English ships present; the English suffered meager casualties (including Nelson). As a result of these successes, after almost two centuries and far beyond their useful life, the *Fighting Instruction*'s grip upon the Royal Navy permanently loosened.

EXPERIENCE-BASED DECISION MAKING

Relying on experience is a powerful methodology for making decisions when we use it properly. By experience, we mean the aggregate of what an individual has learned from the process of dealing with problems and making decisions in the course of his or her life and career. Viewed this way, experience falls into one of two categories: (1) memories of actual events and (2) rules of thumb, judgments, and intuitions that represent the lessons learned from living through those events. Some of these lessons are quite explicit because you can explain what you think about a particular type of problem and why. Other lessons are more subtle or tacit. You may not know exactly why you feel as you do in a given situation, even though you are confident that you know how to deal with it. We often refer to this as intuition or, perhaps, instinct.

As in procedure-based decision making, pattern matching is essential to experience-based decision making. But, rather than relying on recognizing templates learned by rote, we compare the problem to similar problems that we have solved before. If we find a good match, we apply the option that worked previously to solve the current problem. Usually this process takes place very rapidly, often intuitively. If someone asks us why we have made a particular decision on this basis, we may not be able to answer clearly, because we are not fully aware how we sorted

2. Alfred Thayer Mahan, *The Influence of Seapower Upon History 1660-1783*, 12th ed. (Boston: Little, Brown & Co., 1890), pp.265-268.

through our library of experiences to find a match. This type of decision making is most valuable and successful when the decision maker has a broad range of relevant experiences and there is not very much time to make a decision. Logically, we expect that a more experienced decision maker will make better decisions, as long as there is a match between personal experience and the problem.

Numerous studies confirm the relationship between experience and skill. For example, a pilot's skill correlates with his or her experience measured in flight hours. When we need a surgeon, we ask (or should) how many similar procedures he or she has performed. We should not be surprised that we find exactly the same correlation between skill and experience in defense decision making. This means you should give substantial weight to what your experience, and that of others, tells you about how to resolve a particular problem. This includes listening carefully to your intuition and the other subtle forms that valuable experience-based judgments can take.

At the same time, we must exercise care because experience-based decision making can be misleading for several reasons. First, just as when we apply procedures, we may not know whether our experiences are applicable to the current decision. What may seem at first to be a familiar type of problem may turn out to be quite different from anything we have ever experienced, i.e., we may erroneously conclude that we can apply the lessons of the past to the present, although the present may actually be quite different. We compound the error of mis-recognition when we take so much pride in our experiences that we are reluctant to acknowledge that they may not be relevant. Experience, and the judgment stemming from experience, can be a source of self-esteem and authority. We may be reluctant to surrender that authority by acknowledging that a decision is entirely new to us. In these situations, we may be tempted to stretch our experiences to make them fit the current problem. Yet logic tells us that the harder we have to try to make our experiences fit a decision, the greater is the chance that the situation does not mirror our experiences very well. Poor decision making is the result.

Second, we have difficulty accepting that some experiences, once a source of effective decision making, have become obsolete. The half-life of an experience can be short, particularly in a time of rapid technological and international change. For example, combat experience in earlier wars may no longer help us make major modernization and force structure decisions today. Similarly, experience from earlier periods may not help us resolve contemporary personnel issues pertaining to gender, race, operational tempo, child care, spouses' careers, and the like.

Third, even if our experiences are relevant and current, we may distort our memories of those experiences and, therefore, our lessons from them, in significant ways. This is because we perceive our experiences through our five fallible senses. In addition, our memories of events, which may not be accurate even initially, change dramatically over time. Many of the events we remember most clearly did not happen the way we remember them; thus, the conclusions we draw on the basis of these experiences may be faulty. Many careful studies have shown that humans are virtually hard-wired to make certain kinds of errors when we recollect the past. For instance:

- We tend to be overconfident about our memories and the lessons we draw from them.
- We tend to overestimate the importance of the factors we remember most clearly.
- We tend to believe that events occurring at about the same time are probably related to each other—whether they actually were or not.

- We tend to believe that events had to have occurred in the way that they did and, therefore, that they could have been predicted.
- We tend to do a poor job of estimating and using probabilities.
- We tend to be too slow to revise our lessons and, when we do, we tend to change existing lessons incrementally rather than to create entirely new substitutes.
- We tend to distort our recollections of experiences and their lessons depending upon the current context in which we apply them.

THE 1973 ISRAELI COUNTERATTACKS: EXPERIENCE FROZEN IN TIME[3]

On 6 October 1973, Egyptian assault forces crossed the Suez Canal and quickly breached the Israeli defensive Bar-Lev line despite its fifty-foot embankments. Heretofore, the Israelis relied upon the massive height of the barrier and, always strapped for manpower, manned the line itself thinly. The backbone of the Israeli defense was mobile tank and infantry combat groups, supported by tactical airpower, designed to quickly counterattack and eliminate any Egyptian penetrations. Their confidence in this particular defensive doctrine was based on Israeli experience in the previous Arab-Israeli War. In 1967, unsupported armored units had easily penetrated Arab defenses, maneuvered with impunity, and captured droves of prisoners. In a similar manner, Israeli aircraft ranged the 1967 battlefield at will, providing excellent reconnaissance to the ground commanders and destroying much of the opposition in their paths.

By 1973, however, the Egyptian Army had changed dramatically. Its anti-tank and anti-aircraft defenses were now layered and many were portable. Their heavier, longer-range weapons were positioned along the Egyptian bank where they could cover a beachhead across the Suez and the lightest systems, often man-portable, went forward with the infantry in great quantities.

After discovering the Egyptian amphibious assault across the Suez Canal on 6 October, Israeli aircraft attacked Egyptian positions all along the canal using their 1967 tactics. The Egyptians damaged or destroyed over half the attackers; in one raid, over 80 percent of the aircraft were hit. The Israeli Air Force was forced to scale back its operations near the canal. Tactical intelligence gathering suffered as a result as the Egyptians carefully consolidated their positions.

Meanwhile, the Israeli mobile combat groups assembled and counterattacked locally using the dramatic rushes that were so unstoppable in 1967. The Egyptian infantry, schooled in new tactics and confident with its new weapons, ambushed the attackers and destroyed almost all the Israeli armored vehicles. Larger Israeli counterattacks that night and the next day were also devastated. One attack lost 90 percent of its tanks in the first ten minutes. Moreover, the Israeli armor division counterattacking the beachhead lost two-thirds of its tank strength in 48 hours of combat to the new Egyptian defensive tactics.

On 8 October 1973, with the Israeli reserves now mobilized and with three divisions available in the Sinai, but with air reconnaissance still inhibited by the Egyptian air defense umbrella and therefore with limited tactical intelligence, the Israeli Army launched a larger, more prepared counterattack. However, the Egyptians held and the Israelis lost 250 tanks. The Israelis went over to the defensive until they developed combined arms tactics to defeat the new Egyptian capabilities. Consequently, the Egyptians exhausted themselves in several offensives designed to relieve pressure on the Syrians in the Golan Heights.

3. Archer Jones, *The Art of War in the Western World* (New York: Oxford University Press, 1987), pp. 602-604.

To summarize, experience is one of the most important sources of good decision making as long as we are aware of its pitfalls and are humble about the human frailties that we cannot fully escape. How are we to exercise that care? The key is to treat experience and lessons-learned as one source of data or evidence to bring to bear on a decision, along with all other useful information from other sources. How to put this into practice takes us to the last method of decision making.

ANALYSIS-BASED DECISION MAKING

So how do we approach complex, unfamiliar decisions without a pattern to follow? Analytical decision making involves carefully taking a problem apart, collecting and testing the evidence we need to address it, then comparing and selecting an alternative. Analysis-based decision making is generally comprised of the following steps:

- Define the problem and the decision maker's objectives
- Select criteria that capture the most important aspects of the problem
- Identify alternatives for solving the problem
- Evaluate the alternatives using the criteria
- Identify the consequences of each alternative
- Assess the risks and uncertainties entailed by these consequences
- Identify the alternative, within the resources available, which performs best

Although we have identified a series of steps above, decision making must never become a rigid set of techniques or a simple checklist. We must flexibly apply our approach. Too many decisions are unique and take their character from a specific, contemporaneous problem. Some decisions, especially technology-related problems, require detailed and precise information about each alternative before we can make an informed choice. Some decisions can be outlined quickly on a notepad or by closing the door and thinking for an hour. Every now and then, time and resource constraints limit us to just a few broad alternatives and require a quick decision with scant information.

Put another way, a decision may not *exactly* conform to the steps described above, but we should consider each step before discarding any of them. Above all, analysis-based decision making requires that we conscientiously cultivate the intellectual habits of objectivity, explicitness, clarity, sufficiency, and skepticism. Sufficiency means that we determine when we have enough information to decide; to do so, we must identify which information is important and which is not. Skepticism means that throughout decision making, we ask ourselves continuously what we think is true or false and why.

The strength of analysis-based decision making is that it enables decision makers to go beyond the limits of procedure and experience. We have already mentioned some of the mistakes we are prone to make when we assess experiences. Add to that list the difficulties we face trying to solve complex problems, those with many working parts. Our unaided ability to handle problems with just three or four significant factors is usually quite limited. Defense problems take us quickly beyond that. If we are defining a service's force structure requirements to support the current strategy, we must consider many historical, political, and military factors, including regional political stability, crisis response times, and the unified commanders' war plans.

The very act of structuring a problem often provides clarity. Analysis-based decision making allows us to deal with complex problems systematically in a step-wise fashion, with each step made explicit and examined separately as a comprehensible part of the whole. The structuring inherent in analysis-based decision making permits us and others to retrace our steps—an additional precaution against mistakes of various kinds.

During analysis-based decision making, we gather and weigh information to determine what is reliable and useful and what is not. Note that in this discussion we have yet to mention numbers. It is a common misconception that analysis-based decision making is synonymous with quantitative analysis. While numbers are often convenient tools, in many cases they are not appropriate tools. Analysis-based decision making at its heart is a way of critical thinking that is applicable to all kinds of problems, quantitative and qualitative, professional and personal. The good decision maker always allows the character of the problem to drive the specifics of his or her approach and never the other way around. When numbers are useful, we use them. When they are not, we do not. Good analysis incorporates objective and subjective information whose quantity and emphasis vary with each problem.

As with the other methods, analysis-based decision making has important weaknesses. First, the process requires time to gather, assess, and interpret information. The time needed may not be great, but *some* time will always be involved. Therefore, when a decision is needed very quickly, reflexively, as in close combat, the formal analysis-based approach may not be appropriate, even though we use analytical decision making to design the procedures and equipment upon which individuals in time-critical situations rely.

Second, analysis-based decision making requires that key information be available about a decision. Virtually all decisions contain some risk and analysis-based decision making can be very useful for identifying its sources and implications. But what happens if the decision is dominated by risks? Analysis-based decision making is inadequate for these kinds of decisions. The prudent decision maker will instead carefully rely on experience, judgment, intuition, and luck. That said, it is all too easy to jump to the conclusion that a particular defense decision is dominated by risk and to discard analysis-based decision making. In the great majority of cases, there is much information we can gather, estimate, or infer. In other words, analysis-based decision making can still be effective in the presence of considerable uncertainty.

Third, we use analysis-based decision making most appropriately to decide among different alternatives to reach a goal. Analytical decision making is less useful for deciding what goal to seek. This is because most goals involve value judgments. For example, what role should the United States play in the world? How much emphasis should the United States place on military operations in its foreign policy? Should the United States actively promote democracies in the world, even at the cost of economic opportunities? How important should force protection and casualty-avoidance be in operational planning? Should all military jobs be open to women? These are critical questions, and analysis can help to address the underlying issues they raise and demonstrate some of their costs and benefits. But at the point where the consideration of values begins to dominate, analytical decision making becomes artificial. These decisions require that we make moral and ethical judgments. After we have done so and set our goal, we can turn to analysis to help us determine how to best reach it.

There are also weaknesses that have more to do with the decision maker than with the analytical approach per se. Analysis-based decision making is frequently misapplied, or, even if properly applied, poorly executed. Over-reliance on quantitative methods is a good example. Many decisions resist quantitative analysis, although they still can benefit from good qualitative thinking. Esprit de corps may be best evaluated *qualitatively* rather than quantitatively, yet many individuals, even experienced professional analysts, are tempted to impose quantitative analysis on such issues. The results are properly criticized as rigid, misleading, and often silly.

The inflexible use of analysis-based decision making can produce excessively academic and impractical results. For example, there is extensive literature on how to make decisions during crisis by using highly abstract mathematical models—models that are too simple to represent reality and, at the same time, too complex to use when time is scarce. Other problems arise when a decision maker depends on methods so technical as to be incomprehensible to others who may need to be persuaded. Decisions must be explainable to those who have to support or execute them.

A similar difficulty is that analytical decision making can sometimes become disconnected from common sense—shorthand for our accumulated experience, intuition and judgment. Sound decision making and sound common sense are completely compatible, but decision makers can become so focused on the *process* of decision making that they can lose touch with reality. We must always subject decisions to the test of common sense. That a decision may not pass this test need not mean it is incorrect. Common sense may be flat wrong because there may be something new at work, like a leap-ahead operational concept (network-centric warfare), technology (stealth), or weapon system (the airborne laser). But, when a decision seems to defy common sense, we need to investigate. While it may be because the alternative we chose represents a genuine advance over what common sense told us was possible, it may also be because the decision became disconnected from reality.

Finally, analysis-based decision making is most likely to go wrong when we have not tailored it to the problem we are solving. The method of attacking any problem must be driven entirely by the requirements and character of that problem. Analytical decision making loses its way when we forget this basic principle. Remember that analytical decision making is not a substitute for experience, professional judgment, and intuition.

ANALYTIC AND INTUITIVE SOLUTIONS: LIMITS AND OVERLAPS

Analytic and intuitive approaches are often contrasted with each other in terms of their approaches, but, for similar problems, do they really achieve different results? In the 1970s, behavioral scientists researched this question. Studies compared intuitive and analytic solutions to simple problems, such as how many people do we need to assemble for the probability of two having the same birthdays to exceed 50 percent (answer: 23) and to more complex problems where hard data were available, i.e., there was a right answer. Peters, et al.,[4] compiled solutions between two groups that used intuitive and analytic approaches and graphed the results. The members of the intuitive group were half as likely to achieve a perfect answer, but the range and magnitude of their errors was much smaller than that of the analytic problem solvers. In other words, when analysis was done correctly, it was near perfect, but when it was done poorly, it was wildly wrong. Intuition provided a higher number of errors (a more general solution) with less danger of being completely wrong.

4. Peters, J.T., et al, "A Note on Intuitive vs. Analytic Thinking," *Organizational Behavior and Human Performance* 12 (1974): 125-131.

RATIONAL DECISION MAKING

Taken together, experience and analysis are our two most powerful techniques for decision making and their connection is often synergistic. The latter allows us to address areas unfamiliar to us while experience informs analysis. Experience, judgment, and intuition, therefore, should be coupled to analysis-based decision making, each making the others stronger, more useful, and more practical. Rational decision making, the combination of reason (analysis) and experience, is therefore where this discourse leads us. By deciding rationally, we subject experience, professional judgment, intuition, and analysis, along with all other sources of information, to agreed-upon standards of rigor, soundness, and explicitness.

Thus, by advocating rational decision making, we are encouraging you to apply a careful combination of experience-based and analysis-based decision making procedures to solve complex problems, recognizing the strengths and weaknesses of each while remembering that their mix varies with each decision. Typically, the military officer and defense civilian bring their operational experience to decision making while the analyst brings specialized knowledge.

Reconciliation and Execution

The issues we encounter implementing a solution to a force planning problem are largely in the realm of Strategic Leadership, a subject of your Naval War College course in Policy Making and Implementation. We consider some aspects of implementation problems in this course as well, for two reasons. First, we can seldom implement a defense decision without compromises. The word "friction" best summarizes the many reasons for this. Our U.S. defense and military organizations tend to be large, complex, and bureaucratic. Executing a decision in such organizations can require the direction and cooperation of large numbers of fallible human beings with their own preoccupations, desires, capabilities, and interests. Execution also requires resources that are inevitably limited. Finally, execution must be subdivided into individual steps that are parceled out to different organizational sub-units, each requiring coordination with the rest.

This problem of organizational and bureaucratic friction means that seldom can we simply implement what we regard as the "best" solution to a problem. Instead, we usually seek a satisfactory solution that can be implemented effectively rather than a "best" solution that may be implemented poorly or incompletely. For this reason, we must consider what our organization is capable of achieving, which may be limited. This organizational limitation may enter decision making as early as when we choose the alternatives to evaluate or later as a more general constraint treated similarly to a lack of resources. There is an old story about two people at a county fair drawn to the booth advertising a dancing dog. Upon seeing him in action, one remarks that the dog dances badly. The other answers that it is a wonder the dog dances at all. Something similar can be said about large organizations and their execution of program and policy decisions.

The second point about execution flows from the first. Whatever course of action we seek to implement, the cooperation of other organizations is almost certain to be necessary. Equally often, these other organizations will have their own preferences for what course of action to select and how to implement it, and their preferences may not be ours. Sometimes we can overcome this disagreement with a simple order; much more often we must negotiate. In any case, we should prefer to obtain the willing cooperation of others rather than to compel it. The process of obtaining that cooperation means we seek solutions to problems that are acceptable not only to ourselves but for the others whose help we need. In this course, we call that process reconciliation. Finding a way to satisfactorily accommodate the interests of many within a particular course of action is the basis of reconciliation. Analytical decision making is a powerful approach for identifying which courses of action will accomplish that better than others. Equally important, the orderliness, openness, and objectivity of the approach can help establish trust and credibility between contentious parties with conflicting points of view.

An Executive Decision Making Framework

As we have discussed, a major strength of analysis-based decision making is that it is an orderly, step-wise, and explicit process. You may find many prescriptions for how to carry out this process in the literature on management, decision making, policy analysis, and economics. These prescriptions are called by different names and their details may differ in minor ways, but at their core they are all basically the same. This is not surprising because the same fundamental rules of deductive and inductive logic lie at the root of all these approaches to decision making. The better frameworks all provide a structured way of discriminating what is important from what is not, of bounding decisions, of blending objective and intuitive factors, and they provide a logical sequence of steps that produce consistent, high quality, rational results.

The approach we use to encourage rational decision making in this course is called an "Executive Decision Making Framework," shown in figure 1-1, and in most respects it is typical of the approaches you will find in academic literature about decision making. The one notable difference is that we incorporate Reconciliation as a distinct step within the framework rather than moving from the Decision directly to Execution. We included Reconciliation to reflect our strong view that we cannot consider the problem solved until we have an option that can be executed. In the Department of Defense, that means some form of acceptance by most if not all the interested parties or stakeholders. The most brilliant solutions are worthless if we cannot execute them or if the costs of executing them—including those due to obstructionism—exceed their benefits.

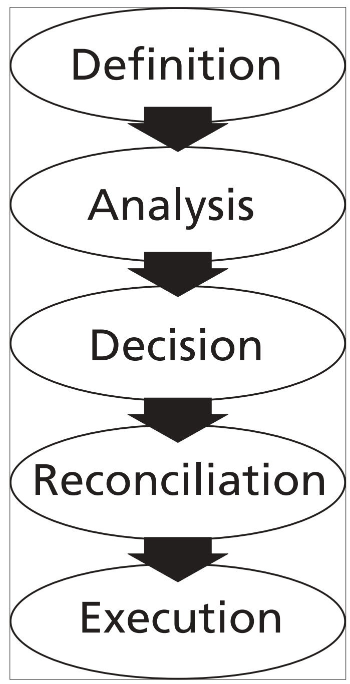


Figure 1-1. An Executive Decision Making Framework

Our Executive Decision Making Framework consists of five steps. This course is devoted to explicating the first four: Definition, Analysis, Decision, and Reconciliation. The boundary between the U.S. Naval War College's Executive Decision Making and the Policy Making and Implementation courses lies between Reconciliation and Execution; addressing Execution, as we mentioned earlier, is the province of the Policy Making and Implementation faculty.

Our Executive Decision Making Framework is *not* a checklist we follow ad nauseam for every problem. As the previous discussion emphasized, the art and science of rational decision making involves tailoring one's approach—the use of analysis and the blend of experience—to the character and the context of the problem. This means that some aspects of the framework will increase or decrease in importance depending upon what we need to do. Therefore, consider the Executive Decision Making Framework as broad guidance on how to organize decision making to ensure we consider each step appropriately.

The Definition Phase refers to the process of fully understanding and specifying the problem we must solve. This step determines the path of subsequent decision making and, most importantly, it enables us to know whether we have solved the problem. In the absence of proper problem definition, everything that follows lacks a sense of direction. The likeliest result is that we arrive at the end of our decision making having lost touch with the problem we set out to solve.

The next step, the Analysis Phase, is multifaceted. It includes developing the criteria we will use to compare the alternatives, selecting the analytical method to make the comparison, evaluating risk and uncertainty, identifying or constructing alternatives, organizing the criteria into a model, and finally using the model to expose the strengths and weaknesses of the alternatives.

We give the Decision Phase its own step to reflect that using the results of our analysis to support decision making requires additional skills. Rarely can we adopt raw analytical results without additional refinements to incorporate the practical considerations of politics, timing, personalities, ethics, spillover effects, and the like.

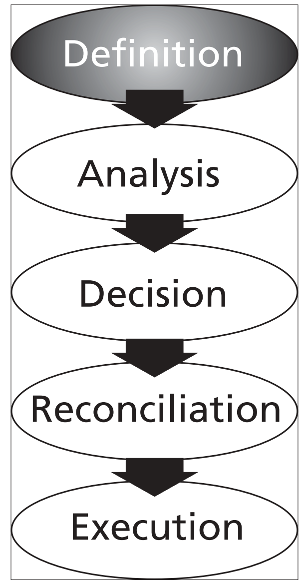
The Reconciliation Phase is the step in which our decision collides with those taken or preferred by the other stakeholders of the problem. In this step, our challenge is to find a course of action that allows us to secure our objectives and that allows others to achieve theirs. Such common ground is the basis for negotiated settlements.

You will find a single page version of our framework in Appendix 1 and an expanded version in Appendix 2. They will help you recapture the principal elements of the following chapters throughout this trimester and later while engaged in making defense decisions yourself.

Summary

With this book, our goal is to couple the military judgment of the professional officer and career defense civilian with the powerful tools of analysis to encourage rational decision making. By doing so, we provide the formal process of defense resource allocation with its most important constituent element: executive decision makers who make the best defense decisions in support of U.S. national security. We do that by giving you a decision making framework and a thorough grounding in analysis, two invaluable tools to help you do the right thing. You bring the experience.

C H A P T E R **2**



PROBLEM DEFINITION

"Difficulties" is the name given to things it is our business to overcome.
-E.J. King: Address to the graduating class of the U.S. Naval Academy, 19 June 1942

A CLEAR PROBLEM DEFINITION IS THE FIRST, and, perhaps, most important step toward rationally selecting the best alternative. Many dedicated and intelligent individuals have produced elegant solutions for problems other than those they were tasked to solve. Therefore, a good executive decision maker participates in problem definition because this step establishes the goal for everything else that follows and places a premium on professional judgment.

In this chapter, we will discuss the opening phase of our Executive Decision-Making Framework: we will examine techniques to describe defense problems in terms that are meaningful to our organization and our decision maker, examine those problems' contexts and boundaries, and then prepare for the Analysis Phase by specifying its objectives. The components of the Definition Phase are shown in figure 2-1 on the next page; we will explain each element in turn and apply them to an example case at the end of the chapter.

The Decision Maker

One of the first areas we must address in the Definition Phase is who we will identify as our decision maker—our approval authority—for the problem we are going to solve now. In the Department of Defense (DoD), there are many decision makers between, on one hand, the development of a concept and, on the other, equipment procurement or policy execution. Beyond the Pentagon, there are more decision makers in the Executive Branch of the Federal Government who evaluate DoD proposals and forward them (sometimes with modifications) in the President's Budget Request to Congress. Within Congress, there are hundreds of individual decision makers. It is they who fund these DoD-originated proposals—sometimes with their own alterations. Moreover, Congress's decisions may not be permanent from one budget to the next, since even funding for multi-year programs, the most stringent type of government budgeting, can be revised or rescinded. Thus, even at the highest level of government, there is no single, final decision maker.

As senior leaders in DoD, whether officer or civilian, we have a chain of command that provides both a forum and a path for our decision making. It also helps us determine who we will identify (beyond ourselves) as the immediate decision maker and the approval authority for our problem solving efforts. In this course, we will generally take the perspective that we are as-

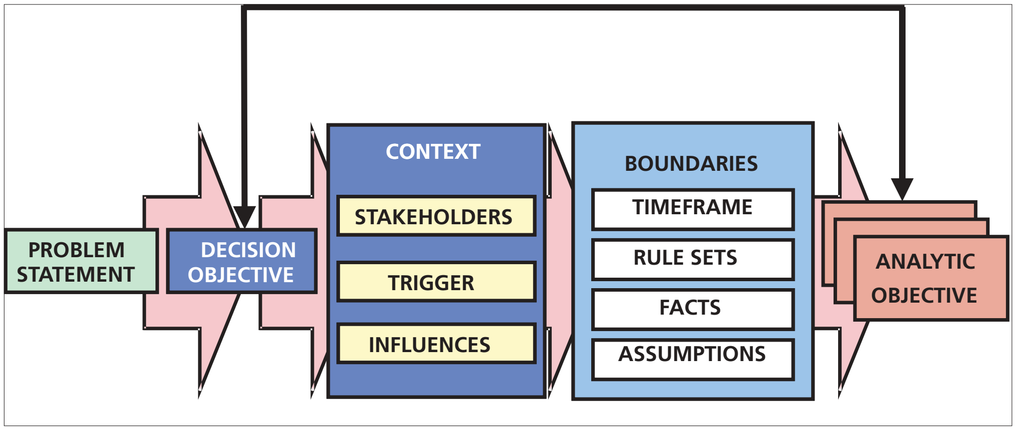


Figure 2–1. The Definition Phase.

signed to a joint staff, service staff, or to the Office of the Secretary of Defense. Our immediate superior tends to participate directly in the work that leads to our decision, whether we are selecting an alternative to promote further up the chain of command or evaluating proposals by others. Occasionally, we may target him or her as the decision maker in terms of our framework, but usually the decision maker we identify is senior to our immediate superior and outside our daily workings. We will therefore assume while using our framework that the decision maker is the first General or Flag Officer in our chain of command. In the Office of the Secretary of Defense or in a Service Secretariat, we may report to a Senior Executive Service civilian. On large joint and service staffs, this is usually the first tier who can approve wide circulation of a concept or approve critical comments on documents that we review from other organizations.[1]

The background and experience of the decision maker may influence how we make presentations, and apply the EDM framework. As their seniority increases, executive decision makers (including ourselves) are less likely to have detailed functional expertise or personal experience to apply to the issue. The decision maker may be a civilian appointee without prior military service. It is sometimes incumbent upon us to educate the decision maker as we describe the problem to ensure that we arrive at mutually understood terms.

As you will see, we highly recommend involving the decision maker at several points in the Definition Phase. In some cases, he or she may provide us with elements of the framework directly, such as the problem statement or the decision objective. In other circumstances, we may want the decision maker to approve our proposals to ensure that we address the issue in terms meaningful to him or her, and therefore to our organization, before we invest significant time and energy executing the remainder of the framework.

Defining the Problem

In general terms, a problem exists when there is a situation that presents doubt, perplexity, or difficulty; or when a question is offered for consideration, discussion, or solution. In the Department of Defense, we have a problem whenever we have a requirement or expectation that is not being or will not be met, whether due to inadequate equipment, organization, doctrine, training or policy. Our recognition that a problem exists is the first step in describing it in meaningful terms. Facile as it may seem, we must ask ourselves whether we really do have a problem. Is there something that we need to fix?

1. Critical comments, in the lexicon of joint staff work, mean that the organization disagrees with a product as written and will object to its continued progress in staffing. The owning office can either modify the product or forward it with the critical comments, realizing that the opposing organization may continue its objections as far as a mutual superior.

As we define the problem, we must determine our expectation or requirement and compare it to our existing and predicted conditions. The difference between what we have and what we need is the magnitude of the problem. The effect of failing to solve it is the problem's importance, which is a value judgment. How quickly this problem needs to be solved is its urgency. Identifying the *magnitude*, the *importance*, and the *urgency* of the problem leads us to decide how many intellectual, physical, and fiscal resources our organization should devote to solving it and how quickly they need to be applied.

ORGANIZING THE PROBLEM

There are two aspects of organizing the problem that are important to us at the beginning of the Definition Phase. The first is whether and how the problem we are solving fits into a larger picture and then how much of that larger picture we have to consider as we proceed. The second is how we want to organize the problem for ourselves and our staff, i.e., whether we want to break our problem into smaller pieces that allow different people to work more or less independently on each.

First, we consider the external aspect; we need to understand the nature of the problem and its backdrop. The simplest case is a stand-alone problem that can or must be considered in isolation. Some complex problems cannot be segmented into smaller pieces or solved with a series of decisions and they do not permit graduated kinds of alternatives. Issues such as whether the United States should retain its unique Marine Corps or whether gays should serve openly in the military cannot be reduced further. This is most often the case with values-associated decisions. While these problems' solutions may involve simple binary choices, the ramifications of a decision may be quite complex.

In DoD, self-contained problems are few and far between for any but trivial problems. In fact, there is a long historical and contemporary list of problems that DoD has treated as if they were isolated problems when they were not, e.g., ship design decisions that affect maintenance activities and training commands; reduced spare parts funding that adversely affects retention and therefore increases recruiting goals. In most cases we will not expand our problem to include these second-order effects, but we need to be aware of them as we proceed. The mapping techniques we are about to discuss can help us clarify the relationship of our problem to other problems.

Second, after considering its external connections, we can use the same mapping techniques to organize our problem internally for the analysis phase, i.e., to decide whether it has more than one moving part and if so how they are connected. Usually, defense problems are complex and multi-faceted. To reduce their complexity, we can organize them into constituent sets of smaller problems in one of three ways: by hierarchy, by linkages, or by sequence.

Hierarchical Problem Structure

To organize a problem hierarchically, we identify the problem at its broadest level and descend into greater level of detail. For example, the U.S. Army conducts a biennial Total Army Analysis to identify its optimal force structure and therefore its end strength or man-power levels. In doing so, the Army categorizes forces in several ways: (1) by the status and availability of soldiers, i.e., whether they are in the Active Component or Reserve Component (the Army Reserve and National Guard); (2) by their purpose as combat, combat support, or combat service support units; and (3) by their missions and likely employment. Different organizations within the

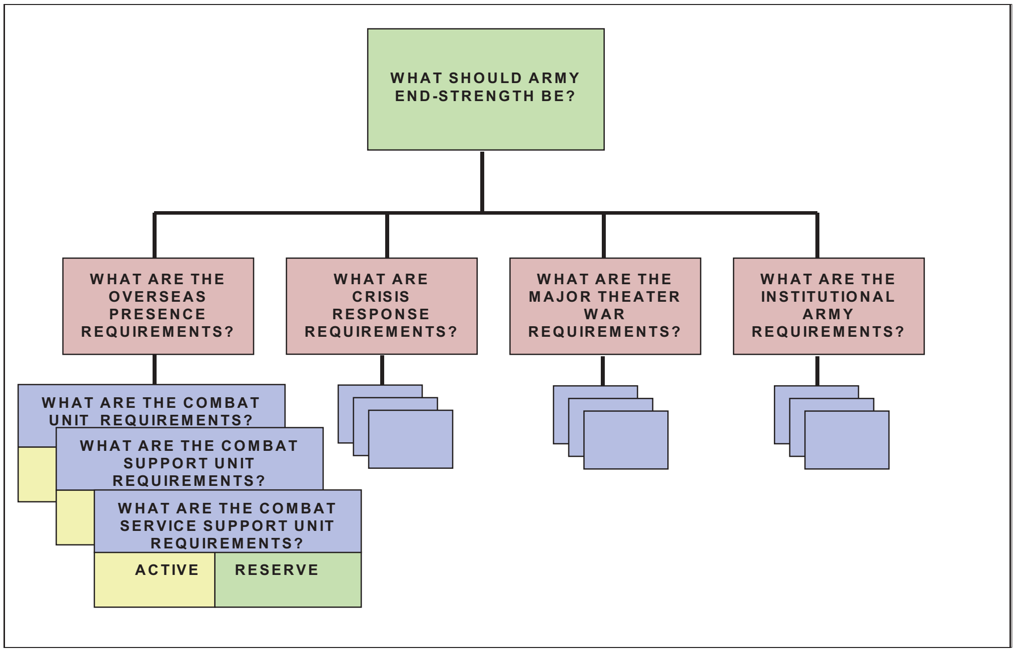


Figure 2–2. Hierarchal Problem Organization

Army emphasize different categories in their planning and there are many claimants for similar units. Army Headquarters determines Army end strength by breaking the problem into components to facilitate its examination and solution.

Figure 2-2 shows one way that Army Headquarters might organize this problem to identify its force structure requirements. They could break it into solvable pieces by sorting force structure requirements first by mission, then by the type of units required, and then by how quickly they are required. After they identify the ideal requirements for each piece, the planners can combine the force requirements to decide where to accept overlaps and gaps. Often, in a situation such as this, the higher headquarters tasks another organization to solve a part of the problem independently. During the Total Army Analysis, Army planners rely heavily on the warfighting commanders of the Central and Pacific Commands to identify forces and their required delivery dates for the major theater wars.

With complex problems that we can arrange into a hierarchy, naturally we would like to start with the broadest problem and then solve its descendents. The inherent danger in this, however, is the sub-optimization that occurs when the broad problem cannot be solved before the descendents. When we recombine the solutions, the compromises necessary to build a composite alternative tend to move away from the optimal solutions of the individual sub-problems. Some participants may resist that movement.

Linked Problem Structure

A set of linked problems requires inter-related decisions because the solution to one problem affects the solution of others. If we decide to shift a particular maintenance mission to a reserve component to solve an active component manpower shortfall, then we may exacerbate a readiness problem and increase the recruiting and training challenges with which the reserves are already struggling. Diagrams of linked problems resemble networks.

Often we must consider the solutions to linked problems simultaneously, as shown in figure 2-3, which features design decisions related to a tactical aircraft. For example, aircraft weight is influenced by the choice of engine (its physical characteristics and fuel consumption, etc.), ordnance delivery requirements and, in turn, aircraft weight will affect speed and range. If we decide not to solve linked problems simultaneously, we usually address the spillover effects of our decision while building alternatives in the Analysis Phase or during the Reconciliation Phase of our decision-making framework.

Sequential Problem Structure

Sequential problems depend on the outcome of a preceding decision to frame and reduce uncertainty about the next problem. Often we can take an incremental approach toward a complex problem by making a policy adjustment or funding research, evaluating the results, and then proceeding if the effort is worthwhile. This is the essential philosophy of John Boyd's popular Observe-Orient-Decide-Act loop and the concept behind branches and sequels in military operational planning.

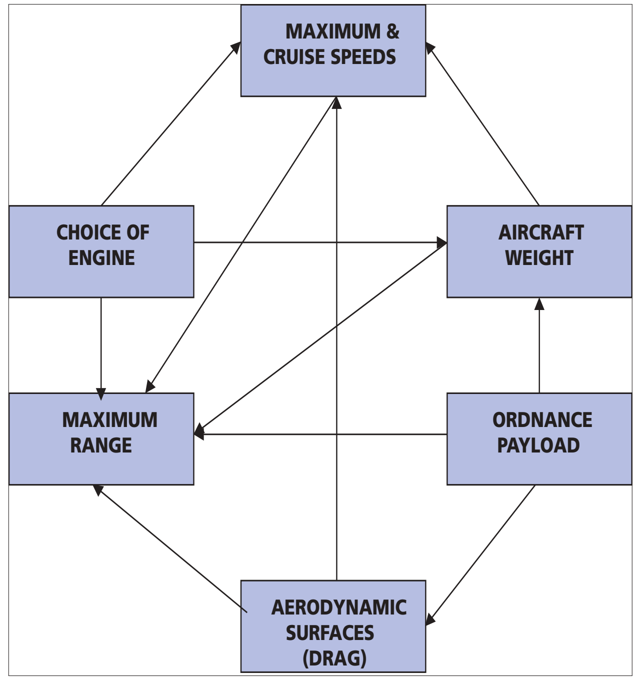


Figure 2-3.Linked Problem Organization

The Defense Acquisition System (DAS), shown in figure 2-4, is a series of sequential decisions (milestones), each of which depends upon the outcome of work that is approved in the preceding milestone. The Department of Defense begins large weapons programs by identifying a need or requirement in general terms and approving its formal exploration at a formal decision hearing or milestone. After refining concept studies and making technology choices, the program reaches another milestone decision and DoD must formally approve its progression. The program will face more reviews and another milestone before it goes into production; each decision point is an opportunity for adjustment or cancellation.

Sequential decision making is a conservative approach that minimizes risk, allows some present uncertainties to be resolved before the next decision, and supports consensus building because the changes from the status quo are neither dramatic nor very threatening. For example, the military is currently running several pilot programs in which private contractors build or manage the family housing on or near a base. DoD's eventual goal is to privatize large amounts of its housing and thereby eliminate the enormous maintenance and repair backlog of government-owned housing. How quickly DoD moves toward privatization, and the form it takes, depends upon the success of the pilot programs. There will be a series of sequential decisions about how to improve contractor performance, by how much, and when to transfer additional military family housing to contractors.

After we understand the backdrop to our problem, we decide which part of the overall problem we are going to solve immediately. The granularity of the problem and the seniority of our decision maker are related to the level of detail at which we

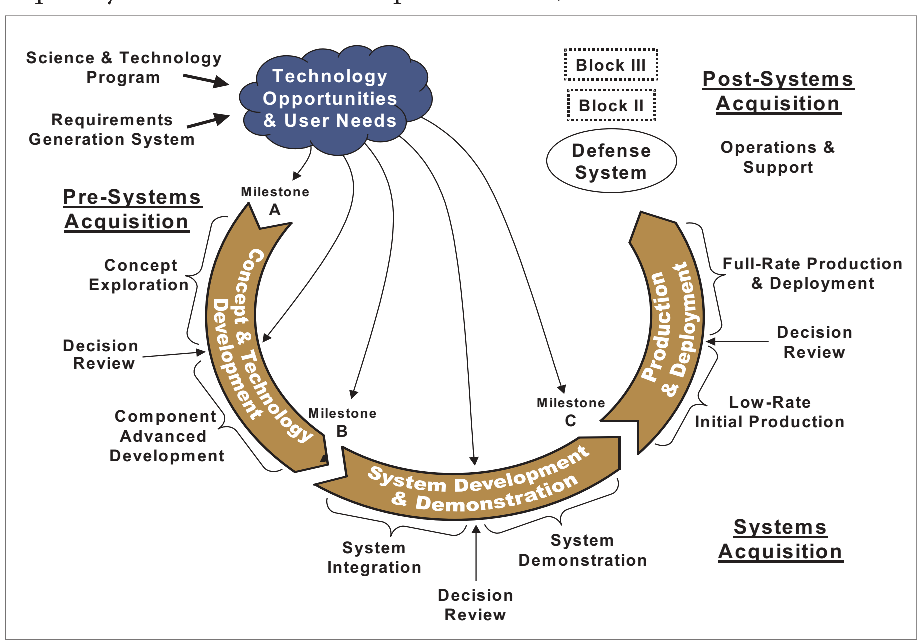


Figure 2–4. Sequential Problem Organization

will solve it. Broad issues suggest broad solutions. For instance, the Secretary of Defense should not decide how many bombs the Joint Strike Fighter needs to carry. Senior leaders should provide overall general guidance and leave it to their subordinates to work through the more detailed levels. During World War II, General Patton prohibited his subordinates from displaying units on their maps more than two levels below their level of command to prevent them from micro-managing and to keep them focused at the appropriate level of operations.

Organizing the problem and deciding at which level of detail we want to address it also impacts our planning horizon for this decision—are we going to solve this problem sweepingly or begin to chip away at it sequentially? Do we want a comprehensive alternative or a quick fix? Naturally, just as more general problems tend to require grander solutions, those solutions are likely to require execution over a longer term. If we choose to present the problem as part of a hierarchy, we expect that the further we descend the more detailed and short-term our problem statements become. Likewise, as we present solutions, we may brief in terms of generalities. Often those generalities are based upon detailed, rigorous problem solving whose intricacies do not interest the decision maker—he or she cares only that they were done and done well.

PROBLEM STATEMENT

We express the results of this process of problem organization as a Problem Statement. The decision maker should approve the problem statement because the single most likely reason for poor decision making is misformulating the problem. Consider the example problem statements below:

- The U.S. Navy needs an operational carrier-based deep strike capability by 2015.
- DoD needs a process to reduce its base infrastructure by 15 percent in the next ten years and it needs a process to identify candidate facilities for closure.
- Some parts of society perceive that the military's gender-integrated basic training is inefficient and encourages sexual harassment.

We said that the problem statement should encompass the appropriate level of detail for the organization that is trying to solve the problem. The first example above is too general to permit us to choose among specific munitions, but it could set the stage for concept development projects by space, aircraft, and missile system manufacturers. The manner in which we express the problem statement is also important. In the second example above, we expect to be asked, "15 percent of what - operating costs, number of facilities, acreage...? Why 15 percent?" If we know what we want and why, we can build this knowledge into the problem statement or answer these types of questions with confidence; if not, we need to better define the problem. In the last example above, we have an instance in which we may not think there is a problem, but someone else does, and we may be tasked to determine whether a problem exists, or even to "prove" a negative, the absence of a problem.

Decision Objectives

Our decision objective is the desired outcome of our organization's decision making. It is analogous to a mission statement.[2] We derive our decision objective from our problem statement. The decision objective is our goal; it provides clarity as we explain our decision process to others

2. If we have arranged the problem into several hierarchical, linked or phased decisions, we may have more than one decision objective within the context of the overall problem. We will consider each of those decision objectives in isolation, however, and for simplicity we will refer to a single decision objective throughout the remainder of the text.

and provides direction for accumulating important information about the problem. Decision objectives should be crafted in terms of solving the problem we identified in the problem statement, and they should not be constrained by the information currently available about the problem. There is a one-to-one correspondence between the problem statement and the decision objective; if we break a complex problem into segments, we create a decision objective for each segment of the problem.

The decision objective is a vital point of reference, therefore we state it simply and clearly, and get it approved by the decision maker. Here are two examples of decision objectives:

- Identify a replacement weapon system for the F-117 Stealth Fighter.
- Determine the least costly method to provide Military Family Housing in the continental United States that equals or exceeds present quality standards.

Each of these decision objectives clearly expresses the expectations of the decision maker. In the first case, there is less specificity in the guidance; the staff may look at current aircraft programs, aircraft under development, new concepts, or non-aircraft alternatives. The verbiage deliberately allows for all those possibilities. If the decision maker wanted us to consider only aircraft alternatives, he or she would specify "aircraft" versus "weapon system" after "replacement." If we are not sure, we should ask. The second decision objective is more specific, requiring us to include cost and quality in our decision—consider how much more difficult it would be to find the "best" way to provide Military Family Housing.

Problem Context

Force planning and policy problems seldom exist in isolation and their circumstances vary in urgency, magnitude, and importance. We regard a procurement cost overrun as a lesser problem for a weapon system being used in combat now than we would while procuring the same weapon in peacetime. As we examine the problem context, we categorize the factors surrounding the decision (in our vocabulary for this course) as Stakeholders, Triggers, and Influences. The stakeholders are those who participate in the decision or are affected by the results of the decision. The trigger is the event that necessitates a decision and determines how quickly a decision is needed. Influences are all the other aspects of the problem that matter and those we will consider in the remainder of the framework.

If we overlook significant factors as we study the context of a problem, we may over-simplify and thereby degrade the quality of the Definition Phase and ultimately the decision itself. If we include too many factors, we create needless complexity and waste resources making our decision. Clearly, after considering the problem's magnitude, importance, and urgency, we must strike a balance between the time available for those involved in the decision to consider a multitude of complex factors and our desire for completeness in describing the problem. Deciding which factors we will consider affects the way we will execute the Analysis Phase, especially the levels of abstraction and simplification we accept in our models.

Because we want a comprehensive list of factors affecting our problem, brainstorming is an excellent technique for identifying stakeholders, triggers, and influences. After we are satisfied with our lists of factors, we can label them as internal or external to the decision. The factors we are going to consider within our organization as we make this decision are internal; the external factors are those outside our organization that we will reconcile later or not at all. Later in the

Definition Phase, we will winnow these factors to bound our problem and simplify our analysis. As we identify stakeholders, triggers, and influences, we may list some factors twice or they may overlap with one another while we describe the problem context. In this portion of the Definition Phase, we concentrate on listing every important factor; precise labeling is truly of secondary importance.

STAKEHOLDERS

In DoD, there are usually a variety of organizations and individuals that are affected by our decisions and, logically, each wants to affect our decision making. In our framework, we refer to any individual or group that has an interest in the outcome of a decision as a stakeholder. Some stakeholders influence us or participate as we make our organization's decision; they are internal stakeholders. Others are external; they may participate later during the Reconciliation Phase or not at all.

Particularly in the Department of Defense, the stakeholders most affected by the results may not participate in the decision making, e.g., changes in pay and housing allowances affect all service members, but very few participate in decisions to adjust them. As we list stakeholders while examining the problem context, we should also list each stakeholder's interest or concern. We want to see whether there are concerns we need to incorporate into our decision, and therefore into the Analysis Phase. We also desire to know how well the other stakeholder's interests align with ours; this will be very important during the Reconciliation Phase. Each stakeholder has its own perspective, offers unique opportunities, and presents certain obstacles toward solving a problem.

TRIGGERS

Each problem that requires a decision bubbles to the surface because of some underlying force, for good or ill. Understanding what compels a solution now, in the near term, or later is an important part of the sorting process used by senior leaders. The trigger is usually what determines the urgency of a decision. Some triggers are highly visible and call immediate attention to a problem, some triggers arise from regularly scheduled events, and others result from good planning. A series of sexual harassment incidents requires some immediate, visible corrective action such as forming an investigatory committee and initiating a policy review. The President's budget must be submitted every year in February and from that deadline DoD reverse engineers its formal resource allocation process with its myriad of decisions.

Triggers, too, may be internal or external; a public outcry is an external trigger while organizations with sound strategic planning impose their own reviews to create internal triggers that lead to important decisions. We identify the trigger so that the decision maker and we become clear on why we are addressing this problem now and to agree on which stakeholders' satisfaction matters most.

INFLUENCES

We define influences as factors we know at the beginning of the problem that affect our decision maker's selection of a procurement or policy option. Influences are background information we are going to consider while making this decision. Internal influences are the concerns we are going to address now while solving the problem within our organization; they are not necessarily under our organization's control, but we are going to factor them into our decision. They will affect our activity in the Analysis Phase. We set aside external influences while making our organization's decision, realizing that we may very well address them during the Reconciliation Phase as we build a consensus on accepting our chosen alternative. Designating an influence as external does not mean we discount it. External influences are germane to the problem and may later come to dominate our senior leaders' choice of which alternative we implement.

Deciding which influences are internal often depends on where our organization sits in the chain of command. Influences that we consider external may be internal to the next decision maker. For example, as the Deputy Secretary of Defense considers the requirement to modernize U.S. tactical aviation for the 21st Century, the historical and projected aviation procurement budgets of the Departments of the Navy and Air Force are internal influences. He needs to consider their historical and projected budgets to assess whether current aviation programs are affordable and to create alternatives if they are not. An external influence for the Deputy Secretary is where the new aircraft may be manufactured—within DoD, this is not of immediate importance to us. Later, during the reconciliation of DoD's proposal with Congress in the federal budget process, the Deputy Secretary will unavoidably address the manufacturers' locations. For our organization (DoD), the manufacturing location is a major external influence, while to Congress it is an internal influence whose importance varies among individual members of Congress.

INFLUENCE DIAGRAMS

A common mechanism for displaying the problem context is the influence diagram. An influence diagram is made up of three principal nodes: decisions (rectangular); random, unknowable, or uncontrollable factors (oval); and evaluations, constants, or calculations (rectangular with rounded corners). We connect the nodes with arrows that indicate how one node influences another. Figure 2-5 is an influence diagram for selecting among shallow water mine detection systems. Performance, cost, and schedule each contribute to the overall value of each prototype. Presently, we do not know the exact characteristics of the mines we may need to detect in the future, which means that our performance evaluation will be based on imperfect information.

While it resembles a flowchart, the influence diagram has several important distinctions. An influence diagram is a snapshot in time, the time of this decision, and therefore it cannot incorporate feedback. Influence diagrams reflect a specific situation, not a process. Also, because they are simplifications, they do not display nuances or important details. They are helpful for diagramming complex decisions and establishing a framework for discus-

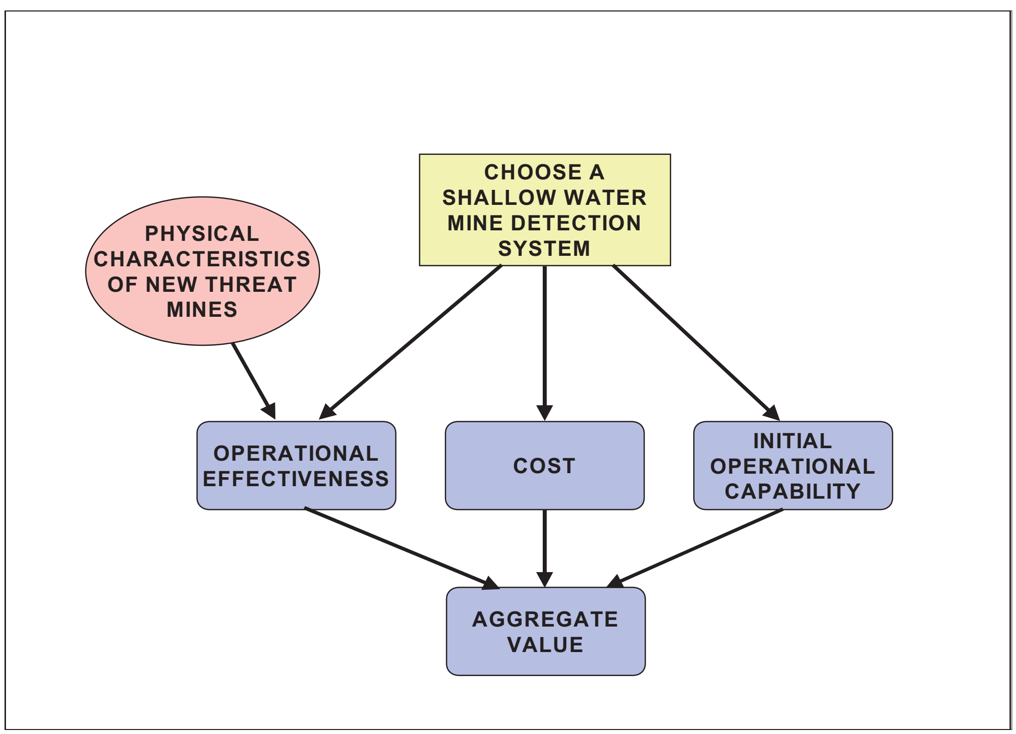


Figure 2–5. Influence Diagram.

sion about context. By building and displaying an influence diagram, participants in a decision create a larger common knowledge basis for the rest of the Definition Phase.

POLITICAL INFLUENCES AND COMPROMISE

We know that the decision maker's background and current position are part of the context of force planning decisions. This is natural in an environment that promotes advocacy and competition between ideas. We also know that what is rational from one perspective may not be so clear from another. Electing to retain a less costly military base in South Carolina and close a similar base in California appears obvious and logical, until we recall that California has already experienced a 50 percent reduction in its DoD facilities during past base closure rounds. Also, many California legislators who will vote on the defense budget were elected (in part) by promising no further base reductions in their state. We may assume that both states' congressional delegations are stakeholders in this decision and that someone is going to be unhappy with the outcome. The politics of the situation may reverse our organization's internally logical decision and recommendation to close the base in California.

We believe, fundamentally, that there is a best policy or procurement alternative for each national security problem that we can and should identify without being influenced by politics. Once we identify our preferred solution, it becomes the "right" thing for our organization to advocate. We need to know what this solution is before we start to reconcile differences with other stakeholders; compromise is often necessary, but we should always know when and what we are compromising, and how far we have moved away from the optimal choice.

Advocacy is an important aspect of American government and is integral to the diffusion of power by checks and balances. The compromises we make, or are imposed upon us, are required by the politics of the situation and are not necessarily shameful. Each compromise should, however, be recognizable by our organization and not be a blind retreat. By knowing our organization's optimal alternative, we may be able to move closer to that alternative later in the process as circumstances change, making some of our retreats temporary rather than permanent. We must also emphasize that people who disagree with our choice are not necessarily wrong or venal. From their perspective as advocates of other organizations, their positions may make perfect sense. The Congressional delegation from California was elected to represent the interests of Californians, which includes jobs and national security. The higher cost to DoD of the California base contrasts with the cost of losing the base to the California delegation; both interests are logical from the perspective of each group.

Politics (in this framework) is generally an external influence in our organization's decision making. One of the reasons we look at external influences during the Definition Phase is to help us prepare for the Reconciliation Phase. We must be able to explain our decision to other stakeholders, some of whom will not like our choice. These preparations may require that we do some additional analysis that may not be necessary for our internal decision, but will help us address the questions we anticipate, the concerns, and the interests of others. In summary, we recognize that we often make defense decisions in a highly charged political environment. We advocate making a rational decision optimized from our organization's perspective and then embracing the political factors to achieve acceptance, and compromise where necessary.

WORKING GROUPS: STAKEHOLDERS AND INFLUENCES

In complex situations that have many variables and components, we may require a multi-disciplinary group effort to define the problem. This is often DoD's impetus for creating organizations, such as the Joint Requirements Oversight Council's Joint Warfare Capabilities Assessment teams that have members from the Joint Staff (often multiple directorates), the services, the unified commands, and defense agencies. Every organization with an important stake in an issue has the opportunity to raise its concerns and announce its preferences with the others in attendance. Representatives are there to contribute and collect ideas and provide early warning—to their own organizations and to others—when an issue or an alternative is controversial. To describe a complex defense problem, even within our own organization, we combine professional military viewpoints to ensure that we develop a thorough, balanced perspective. Our team may be composed of operators and functional area experts, e.g., logisticians, historians, analysts.

Staff processes often take advantage of working groups. Most of the members, and their organizations, are stakeholders by definition. The personality, experience, and background of their members heavily influence the products of working groups. Each participant has the opportunity to affect the group's position on issues, either to promote an agenda, to protect an interest, or to objectively discern what is best for the larger command as a whole. The discussion within a working group and its deliberations may be internal influences on our organization's decision making, or we may treat them as external influences that provide us with insights for reconciliation by indicating who will support or oppose our alternative.

Problem Boundaries

In addition to actually defining the problem, we use the Definition Phase to bound the problem by identifying constraints and limiting the influences we will consider when making this decision. The problem and the decision objective exist within the overall context that we have already categorized and described in some detail. But a decision maker never has unlimited time, personnel, or funding to apply to any problem. To accommodate these restrictions, we limit the scope of the upcoming Analysis Phase by establishing boundaries. We identify boundaries carefully; a poor choice may inadvertently forestall complete analysis, limit the range of alternatives, pre-select one alternative, or introduce bias into the decision. A good choice enables the analysts to efficiently support the decision. We divide boundaries into four categories: Timeframes, Rule Sets, Facts, and Assumptions. We will draw many of our boundaries from our earlier analysis of the problem context, particularly the internal influences.

TIMEFRAME

We have timeframe boundaries of two kinds: deadlines (how long do we have to make this decision?) and planning horizons (how long will we need to implement our solution and how long it will be in effect?) As we identify timeframe boundaries, we return to the problem context, especially our knowledge of what triggered this decision. Importantly, how much time do we have for analysis before we need a decision? One of the criticisms of the Quadrennial Defense Review mandated by Congress is that its deadline of September 1st falls soon after the new administration takes office. Congress should not expect highly detailed force structure decisions because there is simply not enough time to do the analysis and evaluation necessary to inform executive decision makers as they select the particulars of their defense strategy.

The planning horizon is another important boundary. Do we need a short-term solution or a set of alternatives that will solve this problem permanently? Shorter timeframes limit our alternatives to those that show immediate results and therefore favor improving existing systems versus new equipment or methodologies. Similarly, shorter planning horizons lead us toward conservative, incremental approaches. A fund surplus that must be obligated in this fiscal quarter should probably not be allocated to research and development. Long-term planning horizons permit more innovative solutions. Thus, selecting the planning horizon boundary will have an important effect later on how we shape alternatives and how we decide to calculate cost.

RULE SETS

Superiors in our chain of command and organizational culture may establish boundaries that channel our range of alternatives. Also known as value networks, the nature and assumptions behind our present force structure, and its success, have inertia of their own. On the positive side, preserving service culture protects the lessons our predecessors have paid for in blood and treasure. On its negative side, military conservatism or parochialism denies the fruits of technology and growth to those who must execute the next missions. We do not wish to change a service role or culture lightly, therefore one of the primary things critical thinking can help us achieve is to discriminate between truly promising alternatives and those that are merely different.

Closely related to service cultures and roles are the business distinctions between sustaining innovations (existing product improvement) and disruptive technologies (new ways of doing things). The military analogy is whether to execute our current doctrine more efficiently or to adopt new operational concepts, often based on new technologies or systems that may not be mature. We need to know, as we frame our decision, whether our decision maker is willing to consider disruptive alternatives if these represent the optimal solution. We need to know when there are rules that constrain our decision making, either to abide by them or to challenge them.

FACTS

Facts, in our framework, are known truths or "givens" that are not debatable within our organization. Facts may be truths from the historical or physical realms, performance thresholds or objectives, cost limitations, timelines, or any precondition that affects the range of viable alternatives. For example, we may treat Key Performance Parameters for weapons systems as facts. An airplane must fly x distance and back or it is not an eligible alternative. Aircraft range bounds the selection of alternatives and is not negotiable below a certain level.

Givens are suppositions provided by a higher headquarters that we regard as unassailable, although they might not be proven, e.g., we require a force structure capable of executing two overlapping major theater wars. We derive many of our facts (and assumptions) from the lists of influences we made earlier describing the problem context. An example of an internal influence we may classify as a fact is a legal restriction that limits procurement alternatives to U.S. manufactured equipment. We document our facts to keep our decision structured and to display our thought process to the decision maker. If he or she disagrees with our selection, we prefer to adjust the list, before proceeding into the Analysis Phase.

ASSUMPTIONS

Assumptions are suppositions we make in order to proceed with decision making; others may challenge them within our organization and later during reconciliation. They are statements we take to be true without proof, and therefore we should limit them to the absolute minimum necessary to proceed into problem analysis.

Assumptions help us bound our problem and we often use them to place the alternatives on a level playing field to simplify comparisons. We may make an assumption about the projected rate of inflation for the next ten years when we tell contractors to provide us with total ownership costs for their proposals, or by fixing a student population to size training facilities. In defense planning, we make assumptions about enemy capabilities and when they will be operational. The number of assumptions we find acceptable in a problem definition is a function of the problem's importance, magnitude, and urgency.

Poor or hidden assumptions may create fatal flaws in the quality of analysis that affect the decision. One of the reasons many defense experts received Defense Secretary Aspin's 1992 Bottom-Up Review coldly was its easily questioned assumptions, e.g., any and all lesser conflicts could be executed by the two major theater war force structure.

When an assumption is necessary to proceed, but we cannot be certain of or agree to its assigned value, we may specify that analysts use a process called sensitivity analysis. With sensitivity analysis, we explore changes to the assumption—such as the use of weapons of mass destruction by an enemy force—and assess the effect of the changes on the outcome of our decision (see Chapter 7 for more detail). By minimizing assumptions and using sensitivity analysis, we seek to diffuse controversy about the problem definition and decision process and direct the debate toward outcomes and choices.

As with facts, we document our assumptions and have the decision maker approve them. Because assumptions are more subjective than facts, the decision maker's approval is more important for assumptions. If he or she disagrees with them, we must modify them now, before we begin the Analysis Phase, or we may put the entire decision at risk. Our assumptions should appear early in the reports that record our decision making. We, with our analysts, will probably have to make additional assumptions during the Analysis Phase.

Analytic Objectives

Most defense decisions require supporting analysis to ensure that we choose rationally; the analytic objectives are our bridge from the Definition Phase to the Analysis Phase. Analytic objectives are influenced by the problem context and the problem boundaries and may be derived from them directly. The analytic objectives must clearly support the decision objective. If the problem context and boundaries lead to analytic objectives that seem disconnected from the decision objective, something is wrong; we need to review the latter or reexamine the context and boundaries.

Organizations in DoD often contract professional analysts for decision support analysis. These analysts have varied backgrounds that may or may not include military experience. Defining the objectives for our analyses requires a large injection of informed military judgment, i.e., we dare not leave the analysts to their own devices.

In our framework, we always have at least one analytic objective that is subordinate to the decision objective. In even the simplest of cases, we separate the analytic objective from the decision objective because *analysis alone does not decide the issue for us;* it is a tool that we couple with professional judgment to achieve the decision objective. With increasingly complex decisions, we often have multiple supporting analytic objectives. When there is only one analytic objective, its phrasing may be similar to that of the decision objective; however, because analytic objectives describe the goals of the supporting analysis, they often begin with words like "compare" or "evaluate."

Our analytic objectives must lend themselves to independent study; the analyst, with our help, must be able to isolate this subject for study. For example, if our decision objective is to select a new medium-weight truck for the Army, our supporting analytic objective might be: Compare the manufacturers' proposals on the basis of cost and effectiveness for medium-weight truck fleets of 5,000, 10,000, and 20,000 vehicles.

Similarly, each supporting research effort or study should have its own analytic objective. We may arrange related analytic objectives hierarchically or sequentially to indicate when one analysis must precede another. In the case of the Military Family Housing decision objective we used earlier, we may establish the following analytic objectives:

- Compare the cost and effectiveness of government-owned and managed housing, privatized housing, and housing on the civilian economy for military families on bases, stations, or posts within a one hour commuting time of a city of 500,000.
- Identify candidate bases, stations, and posts for pilot program conversions to develop more cost-effective Military Family Housing programs.

Our first analytic objective must be completed before we can proceed to the second. Both analytic objectives suggest that we may require subordinate analyses, i.e., we know it is not practical to do a 100 percent survey of all housing sites; therefore, our first sub-objective under the first analytic objective is to "Build a Data Base of Housing Areas" based on location criteria. We link the objectives of each supporting analysis in the hierarchy to an analytic objective that supports the decision objective that we tied directly to the problem.

Summary

The Definition Phase is extremely important because without appropriate guidance and structure, we can spend large amounts of time diligently solving the wrong problems. Active participation in the Definition Phase gives the decision maker considerable influence over the future course of analysis. The Definition Phase therefore involves much more than just identifying a problem or saying we are going to make a choice among alternatives.

At the beginning of the phase, we craft a problem statement, organize the problem, and specify our decision objective. We identify the context surrounding the problem in terms of stakeholders, triggers, and influences, usually by brainstorming to capture every nuance of the problem. Together, these factors provide important background information that helps us further frame the problem. Next, we establish problem boundaries that will refine our effort, mindful that the timeframe, rule sets, facts, and assumptions set the stage for the forthcoming analysis. Finally, we select analytic objectives. Subordinate to the decision objective, they focus our efforts and limit our scope during the Analysis Phase. We involve the decision maker throughout the Definition Phase to ensure that we are solving the problem he or she wants solved before we begin the much more costly Analysis Phase.

CASE STUDY: THE DEFINITION PHASE
USMC MEDIUM-LIFT REQUIREMENTS: THE V-22 OSPREY AND HELICOPTERS

Background. Shortly after World War II, the U.S. Marine Corps became interested in using nascent tilt-rotor technology for troop-carrying aircraft in vertical assaults. A tilt-rotor air-craft takes off and lands like a helicopter but tilts the rotors on its wing tips forward to achieve much faster forward flight than a helicopter. By 1980 Marine Corps doctrine specified one third of an amphibious assault force would land by helicopter beyond the beach in the Amphibious Operations Area and that helicopters would deliver many of the supplies from ship to shore needed by the assault force. Senior defense leaders also knew that the medium-lift helicopter that fulfilled this role, the CH-46, had to be replaced by the end of the decade.

In 1981 DoD created a Joint Tilt-Rotor Program to explore using tilt-rotor aircraft for medium lift and designated the Army as the lead agency with the Marines, Navy, and Air Force all participating. DoD projected it would purchase 913 aircraft. By 1988, however, the Army had withdrawn from the program, the Marines had taken the lead, and the Air Force had reduced its buy for a new projected DoD total of 657 aircraft.

By the end of the 1980's, Marine Corps doctrine embraced over-the-horizon amphibious assaults to reduce the vulnerability of ships off-loading onto a beachhead within sight of the coast. Over-the-horizon assaults required three new weapons systems: the Landing Craft Air Cushioned (hovercraft) that could lift heavy but non-assault loads quickly from ship to shore; an Advanced Amphibious Assault Vehicle for forced-entry surface assault and protected mobility ashore;[3] and higher-speed, longer-ranging, medium airlift for the vertical assault element. This trio of new equipment would allow the Marines to depart further from seaward and to range deeper into the Amphibious Operations Area. The V-22, Bell-Boeing's tilt-rotor, 100 knots faster than comparable helicopters, was the Marines' preferred medium-lift platform.

The Marines believed that the speed and range advantages of the V-22 were so important that they crafted the Joint Service Operational Requirement (the equivalent of the current Operational Requirements Document) to mandate transit speeds of 250-275 knots, speeds that only the V-22 could meet. The V-22 was, however, much more costly than the helicopter alternatives. The medium-lift helicopter fleets, upgraded to V-22 avionics and electronics standards, cost roughly $46M (FY88 constant dollars) each, while the V-22 cost $67M.[4] Within DoD, a controversy arose whether the additional capability of the V-22—unquestioned by all—was worth the cost. In 1989, Secretary of Defense Cheney canceled the V-22 Program, citing near-term costs as one of the most compelling reasons, and he proposed meeting Marine Corps medium-lift requirements with a mix of CH-60 and CH-53E helicopters.

3. The AAAV passed Acqusition Milestone I began Program Definition and Risk Reduction in 1996. Prototype testing began in 2000 and the Marines expect the AAAV to achieve Initial Operational Capability in 2006.
4. Derived from Simmons, L.D. et al, *Assessment of Alternatives for the V-22 Assault Aircraft Program*, Executive Overview, Institute for Defense Analysis, 1991, pp. 11-12.

Several congressional committees were unhappy with Secretary Cheney's decision. In the 1991 defense authorization and appropriations bills, the House and the Senate directed DoD to commission an independent Cost and Operational Effectiveness Analysis of the helicopter and V-22 alternatives. They specified five helicopter options[5] and six missions[6] and, in addition, told DoD to consider the vulnerability and likely combat attrition of each aircraft option. DoD commissioned the Institute for Defense Analysis (IDA), a Federally-Funded Research and Development Center, to conduct the Congressionally-directed study.

With this background, we will use the IDA Medium-Lift Study as a running example throughout the text to apply our Executive Decision-Making Framework as if we were on Secretary Cheney's staff in 1991. The complete five-volume Secret IDA Study, plus Executive Overview, is available in the Naval War College's classified library under call number U390 15 R-371 (S). Appendix 3 of this text contains the 19 July 1990 record of testimony before the Senate Appropriation Committee hearing on the V-22. It begins with a detailed overview of the study by its author, Dr. L. Dean Simmons, followed by a rebuttal by OSD, and some pointed questioning by several senators.

THE DEFINITION PHASE

Problem Statement. DoD, in particular the Marine Corps, needs a medium-lift aircraft to replace the aging CH-46 helicopter fleet.

Decision Objective. Identify the best alternative for meeting DoD's, particularly the Marines', medium-lift requirement.

Problem Context. Secretary of Defense Cheney canceled the V-22 program because he felt that DoD could acquire adequate medium-lift capability for the USMC amphibious assault mission at significantly lower cost by procuring a fleet of helicopters. He took this action on the advice of analysis done within the Office of the Secretary of Defense's Program Analysis and Evaluation Office (OSD/PA&E) then headed by Assistant Secretary of Defense David Chu. This action, Dr. Chu believed, would free much needed funds for other programs with a marginal loss of capabilities in Marine Corps medium-lift.

Others, including the Marine Corps, some members of Congress, and interested defense contractors, argued that Secretary of Defense Cheney and Dr. Chu were wrong and that the V-22 program should be continued. Congress continued to appropriate funds for the V-22. They directed, in the aforementioned bills, that DoD provide an independent study that compared the V-22 and a range of alternative aircraft packages in Marine Corps missions and a variety of other missions as well. All of the participants agreed that a decision was needed urgently to replace the aging medium-air fleet.

The biggest disagreement about the Marine medium-lift replacement aircraft was about magnitude: whether the much more expensive V-22 would provide a revolutionary capability required to execute over-the-horizon assaults or whether the V-22 was gold-plated medium-lift that helicopters could in fact achieve the mission, albeit less elegantly.

5. "... Including, but not limited to, CH-53E, BV-360, EH-101, CH-46E, CH-60 and any combination thereof....." *H.R. 2461*, Report 101-121, July 1, 1989.
6. Amphibious ship-to-shore movement, follow-on operations, long-range Special Operations, Over-The-Horizon landings, drug interdiction, and Combat Search and Rescue. *H.R. 3072*, Report 101-345, November 13, 1989.

STAKEHOLDERS

Legislative Branch

- Congress as a whole is concerned about cost-effective national defense.
- Some Congressmen are concerned about V-22-related manufacturing jobs in their home states and districts.

Executive Branch

- Secretary of Defense Cheney (our decision maker) is concerned about the affordability of all DoD programs and their relative priority to one another.
- The Secretary of the Navy has conflicting interests; he wants to support the Marines yet he needs funding for four other Navy Department aircraft programs: the A-12 Avenger II medium strike aircraft, the new Maritime Patrol Aircraft, the SH-60 helicopter, and the F/A-18 Hornet strike fighter.
- The USMC Commandant is adamant; he wants this aircraft because it is essential to the Marine Corps' future operational concepts.
- The Chief of Naval Operations is concerned about protecting the aforementioned Navy aircraft programs although he would like to buy 50 V-22s for Combat Search and Rescue.
- Other service secretaries and service chiefs have limited interest in the V-22; the Air Force may make small quantity purchases for special operations missions, but it is not a high priority; the Army is indifferent and believes that too much money in general goes to support expensive aviation programs.

Contractors

- Bell/Boeing (now Textron) anticipates at least a $40B program from DoD, and with the research and development already paid for by DoD, a lucrative commercial opportunity to offer civilian tilt-rotors at competitive prices (compared to helicopters).
- Sikorsky and other helicopter manufacturers will gain an important contract if their air-frame is chosen to replace the CH-46.
- Japanese aircraft companies will develop tilt-rotor technology if Bell/Boeing does not; they are currently behind, but with no competition at all they could corner this market.

Labor Unions. The Texas and Pennsylvania factories would employ 2,000 people each through 2014 manufacturing the V-22. Altogether, according to the manufacturer, the V-22 program will sustain 15,000 jobs in 43 states. The United Auto Workers added V-22 funding as a "key vote" to its congressional scorecard, an important tool they use to distribute campaign funds to election candidates.

At this point, from IDA's point of view, all the stakeholders are external to the decision except the Secretary of Defense. They are doing the study for him and, while they will use data from other sources, like the Marine Corps, IDA will not allow them to participate directly in their analysis.

TRIGGER

The Marine's requirement for a wholesale replacement for the aging CH-46 medium-lift fleet is an internal trigger; the demands of Congress and the forthcoming DoD budget sub-missions (external triggers) mandate a decision as soon as possible.

INFLUENCES (E-EXTERNAL; I-INTERNAL)

- Aging CH-46 fleet; we need a decision soon. I
- Competition among major DoD programs for limited funding. I
- Six earlier studies supported V-22 procurement. E
- Amphibious assault medium-lift mission requirements dominate the other scenarios. I
- Marine Corps says V-22 is essential to Over-The-Horizon capability, predecessor to Operational Maneuver From The Sea (the current USMC operational concept). I
- Marine Corps scenarios require a medium-lift aircraft with a 200 NM tactical range. I
- DoD allocated $24B (FY88) for a replacement CH-60/CH-53E fleet. I
- The Marine Corps desires 425 V-22s to be able to lift 50% of the vertical assault force for the initial wave. I
- The Navy still has a requirement for 50 Combat Search and Rescue aircraft. I
- The Air Force still has a requirement for 50 Special Operations aircraft. I
- Distribution of V-22 jobs: 15,000 jobs in 43 states. E
- There is congressional pressure to reduce the defense budget and a presidential mandate to cut $6B from DoD's budget request. E
- V-22 leading-edge technology has vast commercial opportunities for U.S. aerospace industries. E
- Foreign competition: the Japanese were considering starting their own program that would compete in the civil aviation sector. E

BOUNDARIES

TIMEFRAME

The study had to be completed quickly before the next budget submission. Both the Marines and DoD sought a permanent and complete solution to medium-lift; therefore they were looking at a distant planning horizon. They framed the decision of which aircraft to purchase against a thirty-year timeframe, the expected time required to produce the aircraft fleet and its service life.

RULE SETS

The V-22 was in many ways a disruptive technology, but the Marine Corps had already embraced it. DoD had no objection to the technology itself. However, it was very concerned about cost and balancing the outlays for all of its programs in view of budget cuts. In this case, DoD leadership felt strongly that the much greater cost for introducing the new leap in technology that produced a marginal and unnecessary improvement in effectiveness—speed and range—was not justified. Congress, however, employed some additional rules, basically their constituents' issues in terms of jobs, which made the V-22 a very attractive option. The organizational culture of the Marine Corps, Department of Defense leadership, and Congress influenced their decisions regarding the worth of the V-22 compared to helicopters.

FACTS

- Congress specified scenarios for comparisons between the helicopters and the V-22; DoD could add more.
- The V-22 is 100 knots (nautical miles per hour) faster than helicopters.
- Survivability and attrition rates vary among the V-22 and the helicopters.
- Costs before 1990 cannot be recovered.
- DoD budgeted $24B (FY 88) for the Marine medium-lift replacement fleet.

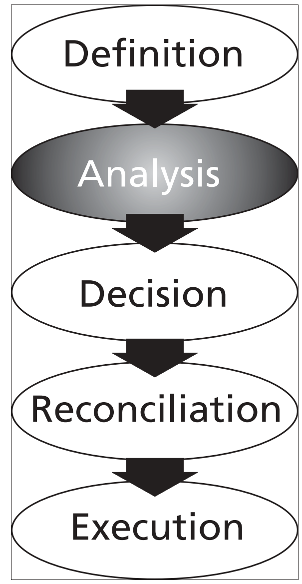
ASSUMPTIONS

- Cold War era force-on-force models are adequate to evaluate assault scenarios (in 1990).
- Contractors' performance specifications and projections are accurate.
- All the helicopter alternatives require V-22 avionics to achieve the mission.
- DoD’s proposed replacement – a modified UH-60/CH-53E fleet – is the lowest cost, minimum effectiveness solution amongst the aircraft alternatives, i.e., there is no less expensive acceptable alternative.

Analytic Objective. Compare the V-22 and helicopter alternatives on the basis of cost and operational effectiveness.

CHAPTER 3

ANALYSIS CONCEPTS: EFFECTIVENESS



Facts are stubborn things; and whatever may be our wishes, our inclinations, or the dictates of our passions, they cannot alter the state of facts and evidence.
John Adams

AFTER WE ARE SATISFIED WITH THE RESULTS OF THE DEFINITION phase, we begin analysis. Our primary goal in the Analysis Phase is to gain sufficient knowledge to meaningfully differentiate among alternatives. Most of the knowledge we seek concerns their effectiveness, their cost, and our uncertainty about the quality of the information we have about each choice. In defense resource allocation, analysis is the coin of the realm; other organizations are unlikely to take our proposals seriously unless we can back them with demonstrably robust analysis. Thus, we will address the standards we use to identify what we want to measure to compare alternatives and their likely consequences.

We may require research before we can begin analysis. In our framework, "research" is collecting original data and taking measurements whereas "analysis" is examining and interpreting data and measurements. We cannot conduct good analysis without sound data; therefore we may choose to be involved with the analyst's methods for data collection as well as with the tools he or she selects for evaluating the data.

Analysis almost invariably requires us to use models to organize our thoughts and evaluations. Models vary from the very simple, e.g., a ratio; to very complex theater warfare simulations, as we will see in the next series of chapters that cover the Analysis Phase. We will begin by addressing the most important constituents of models in the next few chapters, then we will discuss models themselves, and we conclude the phase by demonstrating how models are used in force-on-force and policy analysis.

Action Officers, Decision Makers, and Analysts

As decision makers in DoD organizations, we seldom conduct our own formal analysis of complex problems.[1] Either our in-house analytical unit completes it or we contract with professional analysts. The decision maker is responsible for providing or approving his organization's

1. Major joint and service staffs have resident analysts, usually identified on their organization charts in the J-8 Directorates or as Analysis and Simulations staff assistants to the Commander-in-Chief or Service Chief. Sometimes we execute our own analysis; the Commanding Officer of a unit or base may not have the need or resources to execute a contract for expert analysts.

guidance to the analyst. He often delegates routine oversight to an action officer; indeed the action officer may be the instigator of the ideas that require analysis. Between them, they must provide the analyst with their military judgment, particularly in areas that are intuitive, operational, and experiential. Allowing analysts to proceed without the involvement of our action officers, or without the decision maker's approval of the analytic objectives, greatly increases the risk they will make serious mistakes. If we neglect to provide guidance to analysts, they will create their own, for better or worse.

Our relations with the analyst should be collegial, but we must take his or her background and different perspective into account as we proceed in this phase. Earlier, by our careful construction of analytic objectives in the Definition Phase, we notified the analysts that we are very interested and will be involved in their efforts. Now we seek to combine the powerful mathematical tools of the analyst with the operational experience, judgment, and intuition of military decision makers to sustain our rational approach.

Types of Analysis

One of the first things the analyst and we must agree upon is the kind of analysis that will best achieve each analytic objective. There are three basic types of analysis: Exploratory, Cost-Risk-Effectiveness, and Causal. We decide upon the type of analysis now because it will influence the nature of our modeling later.

EXPLORATORY ANALYSIS

Exploratory analysis examines alternatives that are in the early stages of development. During the mission needs and concept development stages of defense acquisition, we cast a wide net because we are looking for the best of all possible solutions. Because we are forecasting environments and encouraging creative, often non-traditional alternatives, we have a large amount of uncertainty and we do not expect very much detail from exploratory analysis. We must examine our assumptions from the definition phase very carefully, sometimes treating them as variables. The results of exploratory analysis are often controversial, so we must structure these studies clearly and exactly, particularly where we have made key assumptions. We should be able to comfortably explain the logic behind them upon demand.

COST-RISK-EFFECTIVENESS ANALYSIS

Cost-risk-effectiveness analysis is the most common type of analysis in DoD; we use it almost universally to evaluate procurement options. Its purpose is to differentiate among problem-solving alternatives, e.g., to select a design for a major weapons system, to allocate funds among competing program alternatives, or to revise the roles and missions of active and reserve forces. When we execute cost-risk-effectiveness analysis, the problem is usually well defined and bounded, and often the alternatives already exist. Therefore, cost-risk-effectiveness analysis generally takes an engineering or mathematical approach. It, too, is hostage to the worthiness of its assumptions.

CAUSAL ANALYSIS

We use causal analysis to determine why something happened in the past, how a previous action created the state we find in the present, or why actions we take now will create results we desire in the future. Causal analysis—establishing cause and effect—is central to making policy decisions, such as discovering why accident rates have increased, how best to conduct basic training, or whether a pay raise or the provision of more recruiters would best increase the number of new enlistees. We want our analysts to rigorously separate facts from values and conduct causal analysis dispassionately. Our values may have entered the decision process in the Definition Phase, but we do not want the analyst to include his or her own subjective opinions unless we so specify.

Selecting Alternatives

We may know the alternatives for solving the problem before we start decision making or we may develop them during the Analysis Phase. When we have the alternatives in advance, that knowledge may help us select criteria and build models that will best expose whatever important differences exist among them. Foreknowledge of the alternatives also indicates the likely range of values we can expect as we evaluate them, saving time and energy by limiting the scope of our analysis, i.e., if we know there are miniscule differences between certain aspects of the alternatives, we do not need to measure them. Nonetheless, our analysis must still be sufficiently general to accept a new, unforeseen alternative and compare it to the options we already have. Indeed, a pitfall of knowing the alternatives in advance is that we may design our model to emphasize differences between the alternatives although these differences may be trivial to the analytic objective. Worse, we may inadvertently favor one alternative for reasons outside the bounds of the analysis by seeking to emphasize differences. Because of these concerns, we always leave open the possibility of generating the alternatives later in the Analysis Phase.

Our set of alternatives should exhibit the following characteristics:

- Breadth
- Viability
- Neutrality

The alternatives must span the scope of possible solutions of the problem, including the extremes as well as the middle of the range of alternative solutions. Extreme solutions may include disruptive technologies that may have enormous spillover effects on our organization and others; they may require delicate handling and we should discuss them with the decision maker to determine whether they are within the boundaries of this problem's solution set.

When we have a continuum of alternatives, we select representative alternatives that permit study and enable clear choices; e.g., most studies of Overseas Troop Strength add increments of 25,000 soldiers as they build alternatives. They identify the capabilities of each force level so decision makers can see how much capability each increment adds. The actual alternative may not be a multiple of 25,000, but the decision maker will have a clear sense of capabilities after reading the analysis.

Additionally, each alternative we study must be a viable solution and meet our minimum requirements; we will not include throw-aways.[2] If an alternative is unacceptable, we should identify whether it can be improved to meet our standards, e.g., a city may be willing to upgrade, at its own cost, the hotel services at its piers or its mass transit to encourage the Navy to homeport ships there. We must be very careful whenever we dismiss an alternative for not meeting our standards; its proponents may ask us later to justify its exclusion.

2. There is an apocryphal story about Secretary of State Henry Kissinger and President Richard Nixon. During a crisis with the Soviet Union, Secretary Kissinger presented the President with three alternatives. "Mr. President," he said, "first, we may begin global nuclear warfare immediately; second, we may capitulate abjectly. I think we should explore the third option more fully."

We strive to shed bias from our alternatives, therefore we describe each in a similar manner with the same level of detail. We test each neutrally, according to the same standards and under similar conditions. One of the traits we value highly in analysis is its empiricism, the fairness we get by testing options and comparing results in a dispassionate manner. A fair competition among ideas is essential to discovering which is best for solving our problem. Besides, each alternative in defense resource allocation will have its proponents, many of whom we will encounter in the Reconciliation Phase and our analysis must be persuasive in that phase; it can only be persuasive if it is thorough and unbiased.

WHEELS VICE TRACKS: THE ARMY'S MEDIUM-WEIGHT COMBAT VEHICLE ALTERNATIVES

When the Chief of Staff of the Army, General Eric Shinseki, unveiled his vision for the Army's transformation to a medium-weight force on October 12, 1999, he was addressing concerns that the Army's heavy forces, although highly capable, were too heavy to move to the fight quickly enough. To reduce the size and weight of the equipment the U.S. Transportation Command would have to lift between and within theaters, he stated that his vision included a new family of wheeled armored vehicles that C-130 intratheater lift aircraft could haul and that would replace tracked vehicles.[3] Most observers understood his desire to lighten up the Army, but it was unclear to many why General Shinseki specified wheeled vehicles in his introductory comments. A variety of senior Army leaders has since said that a family of wheeled vehicles was one likely expression of the Chief's vision and that his comments should not be taken so literally as to exclude the possibility of a new tracked family of medium-weight armored vehicles; all options were on the table and because a wheeled option would break with tradition the Chief chose to emphasize it.

During the following winter, the Army held a vehicle competition for nine contractors with 35 different systems. The only U.S. manufacturer of the three that submitted tracked alternatives was United Defense LP; they introduced reengined, modernized variants of their venerable M-113 armored personnel carrier. Following the demonstration, the Army revised its draft Request For Proposals with some lower performance standards to reflect what they had observed in the trials and to encourage as many contactors as possible to continue participating. United Defense accused the Army of relaxing its requirements because it realized wheeled vehicles could not meet the performance standards while tracked vehicles could.[4]

Army officials denied any bias, but skeptics could not help but note the Army had already leased 46 wheeled light armored vehicles from Canada for use by its two interim brigades as they test new operational concepts central to Army transformation. Senior Army leadership again denied that they had ruled out tracked vehicles, but many of their briefing materials gave exactly that impression in the Spring of 2000. (At an Association of the U.S. Army meeting during 16-19 February 2000, the U.S. Army Training and Doctrine Command's organizational graphics for the battalions of the interim brigades used the symbol for motorized infantry (wheeled vehicles) vice

3. For example, the current M-1A2 Abrams tank weighs 70 tons and can be carried one at a time only by strategic lift aircraft like the C-5 and C-17. It is too heavy for most bridges and maneuvers with difficulty in congested terrain and on narrow roads. By comparison, the maximum weight for the new vehicles is 19 tons.
4. Sean D. Taylor, "Wheels Vs. Tracks: Is Shinseki Moving Too Far, Too Fast?" *Army Times,* 28 Feb., 2000: 12.

mechanized infantry (tracked vehicles)[5] The Army has since created a new symbol that combines both.)

Inspired by concerns that the Army was moving too rapidly toward unproven capability—and at least in part by Congressmen from districts who manufactured track vehicles—Congress held hearings to explore how the Army was selecting these interim vehicles. As a result, while Congress funded the medium armored vehicle procurement program for fiscal year 2002, the Army, despites its protests, was made to hold side-by-side tests of the leased wheeled vehicles against M-113s before full production could begin.

The Army received 20 vehicle proposals on 6 June 2000. Over the next four months, they evaluated 17 of them and tested samples at Aberdeen Proving Grounds using the parameters from their Operational Requirements Document (largely parallel to the revised Request For Proposals that the Army issued the prior spring). After receiving final proposals on 6 October 2000, the Army awarded the contract for the new family of wheeled vehicles to a consortium of General Dynamics and General Motors of Canada on 8 November 2000 based on trade-offs in the following areas.

- Suitability to support operations with the new Interim Combat Teams
- Transportability requirements
- Quality of the training support package
- Technical requirements for the different variants, e.g., characteristics of armament
- Crew protection.

United Defense LP objected to the award decision soon after it was announced, based (they said) on the Army's failure to consider its own requirements, i.e., United Defense LP contends their vehicle is 50% less costly, can be delivered sooner, and that it meets all the Army's performance specifications unlike the wheeled vehicle selected. They also assert that the Army's communications with the contract-winning General Motors of Canada/General Dynamics consortium "substantially exceeded the nature and extent of information conveyed to United Defense LP" and that the competition was pro forma, evidenced by briefing slides prepared before November that incorporated substitute vehicles for the Mobile Gun System variant. (The substitution is significant, posits United Defense LP, because it means that the variant would not be available as required by the timeline specified in the Operational Requirements Document, i.e., the Army was pre-approving a deviation before the contract was awarded… and the delivery schedule was one of the criteria for selecting among manufacturers.)[6]

Conversely, the Army contends the wheeled vehicle family provides overall superior performance, according to its weighted criteria, than United Defense's reworked M-113s. Because of the protest to the General Accounting Office, the Army issued a stop work order on 5 December 2000 while the General Accounting Office reviewed the award.

The Army leadership stood by its decision. In December 2000, Secretary Caldera stated that he believed the selection process would stand up to the Government Accounting Office review

5. Ibid.

6. Kim Burger, "UDLP Offers Additional Evidence of Army Bias in Favor of LAV III," Inside the Army, 15 Jan. 2001:1. UDLP contends the wheeled vehicles failed to meet performance requirements for ammunition storage of ready rounds, separation of ammunition and crew, internal noise, braking, and, for the mobile gun system, battlefield sighting indexing requirements, amongst other shortfalls. They further claim many of the required improvements, such as the mortar variant and swim capability, are high risk and that the armor has to be removed to make it C-130-transportable and question how life cycle costs were arrived at without reference studies.

and that the new administration, like Congress, would find the medium armored vehicle program compelling. General Shinseki has called for armor traditionalists, concerned about the lesser firepower and protection of new vehicles, to stifle their dissent and, "If you chose not to get on board, then that's okay, but get out of the way."[7]

How much of this controversy could have been avoided if General Shinseki had not appeared to exclude tracked vehicles from his vision? Many, in addition to United Defense LP, still feel tracked vehicles are viable alternatives for transforming the Army. For major defense decisions with many stakeholders, the range of alternatives must cover the range of possible solutions without the perception of arbitrary exclusions or we may expect those stakeholders to react to protect their equities.

Additionally, the alternatives must be viable in terms of the problem definition and it is the executive decision makers in DoD and their staffs who will approve the standards against which they are measured. How well were the analysts who designed the field trials in the winter of 2000 listening to their decision makers if they established overly demanding performance standards? The perception they created was that after the trials, when the wheeled vehicles did not fare well, they changed the requirements to make them viable. In reality, the new standards may well be the right ones, but now the issue is clouded.

Finally, there is the issue of neutrality or fairness. General Shinseki let his preference for wheeled vehicles and against the M-113 in particular be known early and clearly to his subordinates. How or whether that affected their decisions we cannot know, but United Defense LP perceived enough bias to raise objections that must be taken seriously—they resulted in the stop procurement order. Seldom can anything good come from promoting a particular alternative without robust analysis to explain rationally why this choice is favored.

The Army leadership may very well have made the right decision to purchase wheeled vehicles. The Light Armored Vehicle III family may best achieve Army transformation goals and therefore be best for the long-term health of the Army. But senior leaders' actions before the formal decision process engaged—especially the Analysis Phase—invited emotional responses. Were tracked vehicles effectively excluded from the beginning? Were the standards changed to ensure a wheeled vehicle choice? With better preparation before the decision and robust analysis, these questions can be answered to the satisfaction (if not the desires) of all the stakeholders during reconciliation; without them, expect controversy and disharmony.

As we consider or build alternatives, we know that the program or policy that is executed after a decision may not be, in the literal sense, any one of the alternatives we constructed in the Analysis Phase. Alternatives may be modified or even combined after analysis to incorporate the strengths of one to compensate for the weakness of another; the executive decision maker is usually in the best position to make these adjustments and usually sees their impact more clearly than the analyst. For example, an auxiliary airfield might be time-shared with civil aviation to defray costs and provide military access to a longer runway than is otherwise available or affordable. Extensive alterations to the alternatives may require that we conduct another analysis.

7. Thomas E. Ricks and Roberto Suro, "The Wheels Turn in Army Strategy: Transformation to Cut Tanks' Role" *Washington Post*, 16 Nov., 2000:1.

Where policy decisions are concerned, we recognize that the more steps or phases an alternative has, the less likely it is to be executed as the originator intended. Each succeeding phase of implementation is actually an opportunity to modify and shape the alternative further. Alternatives that are not phased have an all-or-nothing character to them—often the case in procurement decisions—and they present greater risk of failure for an organization committing to them.

Attributes, Criteria, and Measures

Now we will examine the characteristics of alternatives and decide which to evaluate. Consider we are facing a decision, we have three alternatives that may or may not be effective, and our task is to select the one that best accomplishes our goal. We will do this by predicting the consequences of adopting each alternative and comparing those consequences to one another based on a set of standards we choose. Which standards we select are crucial in the Analysis Phase.

We begin our selection by noting that every alternative we encounter, however simple or complex, is or will be composed of attributes, that is, its entire family of qualities, characteristics, and distinctive features. Size, cargo capacity, weight, speed, and availability rate are typical attributes of vehicle alternatives. Equity, happiness, morale, and quality of life are typical attributes concerning policy choices. When we must select among alternatives, some attributes are more important than others because they are more relevant to the way we defined the problem and the way we expressed the decision objective. In our framework we call these more important attributes criteria; they are the standards upon which we will base our judgments and preferences among the alternatives. The most important subsets of criteria (and the ones we will assess) we call *Measures*, many of which we group in two categories, *Effectiveness* and *Cost*. Effectiveness is the ability of an option to achieve an outcome we desire. Cost is the amount and rate at which alternatives consume resources.

Figure 3-1 illustrates that Measures of Cost (MOCs) and Measures of Effectiveness (MOEs) do not constitute the entire universe of criteria. Schedule, risk, equity, and availability of resources are examples of criteria we measure for a procurement or policy selection that are not strict descriptors of cost or effectiveness. Our collection of criteria may include any number of measures, but cost and effectiveness are almost always relevant to defense decisions.

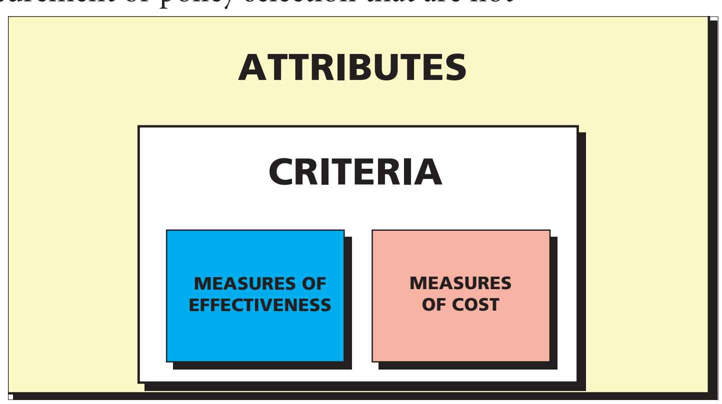


Figure 3-1. Characteristics of Alternatives

When we can evaluate a criterion numerically, we have a quantitative or objective measure. One alternative wheeled armored vehicle has a maximum speed of 65 mph, another 50 mph. A number that we can measure to a very fine degree differentiates them. But we can use numbers and combine them differently to measure alternatives. We do not purchase eggs in the same way that we do armored vehicles. All the eggs in a carton of a dozen do not weigh the same, nor are they the same size. We could measure both egg criteria (size and weight) in every carton we buy, and then buy the carton we find preferable, but there is a simpler way. We buy them based on a qualitative measure; instead of a number we assign eggs to a category that we understand, "large" or "medium," and then choose the carton that will satisfy us. In this case, the qualitative measure (grade) is assigned based on underlying quantitative data; as we will see later, this is not always the case.

Most analysts are more comfortable measuring characteristics that are quantifiable. Left to our own preferences, many of us, too, will gravitate toward a numerical measure; we have less risk of being "wrong" about something we can count. Despite these tendencies, there is nothing inherently superior about quantitative measures, and nothing to suggest that qualitative measures are inferior or less rigorous. The character of the problem must drive the measures we choose.

Selecting Criteria

Our next important task in the Analysis Phase is to identify the attributes we need to measure to support the needs of the decision maker and designate them as criteria. (Later, with the analyst, we will determine how to best measure these criteria.) We select attributes to be criteria based on their ability to indicate whether important differences exist among the alternatives, and, if that is the case, the degree of those differences. Knowingly or not, we generated indications of important criteria during the definition phase. Brainstorming—listing every possibility on a wallboard—is a good beginning for turning those indications into well-defined, candidate criteria. In general, we seek criteria with the following characteristics:

- A direct connection to the analytic objective
- Inclusiveness
- Precision
- Measurability
- Uniqueness
- Discrimination

There are many attributes that distinguish between procurement and policy options that are not germane to the decision and therefore they are not good criteria. Good criteria evaluate the performance of alternatives in the real world in a manner linked to the analytic objective. That is, they help us evaluate the alternatives in a way that matters. For many acquisition programs, the criteria for concept studies are derived largely from the operator's Mission Need Statement that first identified the requirement or deficiency.

We naturally prefer a single inclusive criterion that covers a large portion of the desired analysis to several discrete ones so that we can simplify our data collection and display. We carefully and precisely describe each criterion to eliminate room for interpretation by the analysts or the participants in the decision. We prefer direct, quantifiable measurement to reduce error, even as we understand that such perfection is not always possible.

Each criterion should measure something unique and different from the others. "Double counting"—directly or indirectly measuring the same attribute twice—is usually undesirable, but in exceptional cases may be appropriate. Finally, the criteria should reflect value added for exceeding the minimum requirement to help us discriminate between alternatives. If an option must meet a specific minimum requirement to be eligible for consideration, but there is no value for exceeding that minimum, then that attribute is not a good criterion. It may be an important attribute, a benchmark that each alternative must satisfy, but that importance is not synonymous with being a good criterion. Requirements and thresholds are Go/No-Go filters;

they disqualify an option from further consideration unless the alternative brings itself up to the required standard. Criteria help us compare value beyond minimum requirements.

The more criteria we choose to measure, the more expensive and lengthy the Analysis Phase will be. There is a point of diminishing returns beyond which our attempts to refine the alternatives further are not worth the effort. We may even proceed to the point of over-specification, in which we define so many criteria so tightly that we cannot create any alternative that satisfies them all. Over-specification reduces the effect of individual measures when we weight them in a model. The Definition Phase helped us identify the point of diminishing returns when we evaluated the importance and urgency of this decision to our organization.

After we identify a range of potentially useful criteria, we identify the relative value of each criterion to the decision and determine which criteria we actually want to measure. Ideally, we would like a set of criteria we can measure directly, in quantitative terms. Unfortunately, objective attributes (the quantitative ones) are often far less important than subjective attributes (the qualitative ones). We must guard against choosing criteria that are easy to measure but less relevant to our decision, and we should not shy away from attributes that are difficult to measure.

CRITERIA	MEASURES	EXAMPLES OF HOW WE CAN MEASURE
COST	UNIT COST	CURRENT OR CONSTANT DOLLARS
	PERSONNEL	PAY, MAN-HOURS, MANNING LEVELS
	TOTAL OWNERSHIP COST	CONSTANT DOLLARS
SCHEDULE	FIRST UNIT DELIVERY	CALENDAR DATE
	INITIAL OPERATIONAL CAPABILITY (FIRST UNIT)	CALENDAR DATE OR DATE RELATED TO THREAT
	FULL OPERATIONAL CAPABILITY	CALENDAR DATE OR DATE RELATED TO THREAT
EFFECTIVENESS	MAXIMUM SPEED	MACH, KNOTS, MPH, FEET/SEC
	MAXIMUM RANGE	MILES, KM; EMPTY OR WITH WEAPONS
	WEAPONS LOAD	NUMBER AND VARIETY
	STEALTH	RADAR CROSS SECTION, HEAT SIGNATURE, NOISE LEVEL, SIZE
	SIZE	DIMENSIONS, FT2, FT3, DECK SPOTS, CONTAINER-EQUIVALENTS
	WEIGHT	POUNDS, TONS, DISPLACEMENT
RISK	MATERIAL	% OF COMPOSITES, FIRST APPLIED USE OF MATERIAL
	TECHNOLOGY	NEW OR PROVEN, NUMBER OF TESTS BEFORE PROTOTYPE
	PRODUCTION	NUMBER OF TESTS BEFORE PRODUCTION, % NEW OR UNIQUE COMPONENTS
	POLITICAL SUPPORT	OPERATOR REQUIREMENTS, COMPETING FUNDING REQUIREMENTS, JOB DISTRIBUTION

Table 3-1. Examples of Measuring Criteria.

Assessing Criteria

In the process of identifying and selecting a set of criteria, we must assess the degree to which all of these measurements help us evaluate alternatives that satisfy the decision objective. We examine them, individually and as a group, through three lenses: Validity, Reliability, and Practicality.

VALIDITY

Validity is the degree to which our criteria adequately predict, measure, or illustrate to the decision maker the important differences among alternatives: *Are we measuring the right things to support making this decision?* Are we gathering enough information to make a rational decision? Does each criterion add to our understanding of the alternatives? The set of criteria must somewhere address every aspect of the analytic objectives; when applied to the alternatives, they must help us select. We use analysis to simplify reality; by assessing validity, we ensure that we do not over-simplify or become distracted from the analytic objective.

Put another way, validity is the degree to which we are able to identify what we want to measure. We accept that usually one criterion will not reflect every facet of the alternatives' behavior. There is no single, ultimate criterion we can use to measure the performance of a fighter aircraft. We settle for what we *can* measure: components of the idyllic measure of "fighterness." The most common way to improve validity is to measure more attributes, i.e., to add more criteria, thus, at least in theory, we can move closer to the perfect set of measures that encompasses everything.

A related way to improve validity is to use surrogates for things that are difficult to measure directly. For example, we may estimate aircraft survivability by determining the number of enemy radar types that our electronic counter-measures suite can counter.

On a more abstract level, suppose that we are tasked to evaluate several alternative compositions for U.S. nuclear forces and that the different alternatives' deterrent effect is one of our criteria. This is a tough task, because deterrence is something that happens in the minds of our adversaries (if it happens at all) and is not a directly measurable physical attribute of our nuclear forces. One way to cope with this problem is to use several more directly measurable attributes (e.g., the quantity and size of warheads, their accuracy, and their ability to launch after an enemy attack) as surrogates. If we have reason to believe that our adversaries consider such attributes in deciding whether they are deterred from certain actions, then we can reasonably use these attributes as surrogates for our "deterrent effect" criterion.

The degree of validity for each criterion varies with the problem definition. Consider two decisions about Navy surface combatants. Ships have a huge family of attributes and therefore an equally large set of potential criteria. If we are deciding between builders' proposals to select the design for the next generation destroyer, cost, warfighting effectiveness, technology risk, delivery schedule, habitability, and maintainability are all highly valid criteria. If we are deciding among ships to send on a contingency deployment, a different but overlapping set of criteria is more valid; we may list cost, warfighting effectiveness, level of training, materiel readiness, and command climate as our criteria. Both decisions deal with ships, but they require different criteria; we evaluate each attribute's validity for use as a criterion differently for each decision.

Finally, we check our set of criteria again against our analytic objective to ensure we have not left out an important attribute or that we have not accidentally double-counted the same attribute. We examine each criterion individually to ensure it contributes meaningfully to our understanding of the alternatives.

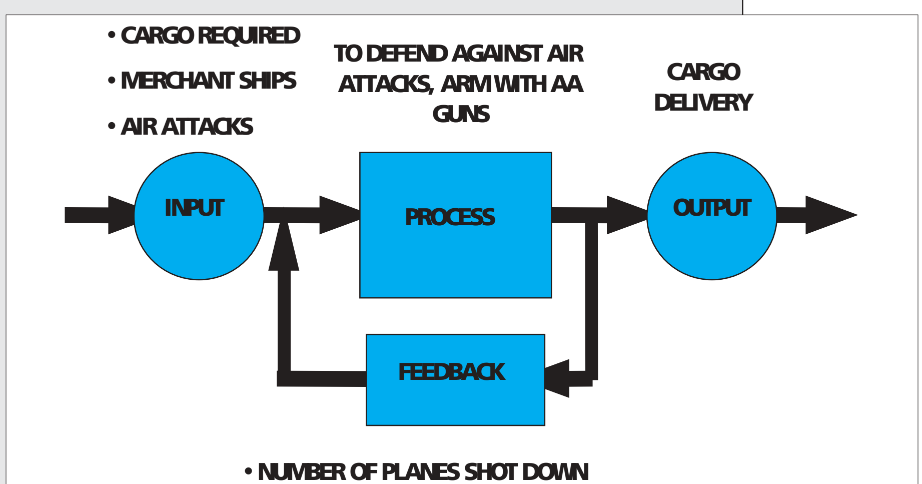
MEASURES OF EFFECTIVENESS: THE SYSTEMS APPROACH AND CONVOY PROTECTION[8]

We often use a Systems Approach as a tool during analysis, consciously or otherwise. In our lexicon, a system is the full array of elements (people, equipment, and processes) that operate together to perform a mission, create a desired state, or achieve an objective. Each system has input and a process that produces an output. We measure a system's output while the system is in operation to modify or adjust the process controls—feedback—to keep us heading toward the goal. The process becomes a loop, not just a linear path from input to output. We use the systems approach in mechanical applications ranging from driving an automobile to guiding missiles. We use similar feedback mechanisms in policy analysis, from efforts as diverse as dieting and exercising, to reducing the national debt and improving student population test scores.

Although widely applied, the systems approach is not applicable for every analysis. If we can easily identify systemic elements in our problem and its potential solutions, such an approach is an easily grasped and appropriate tool for creating simple models of processes. Consider the usefulness of the systems approach as a convenient structure and display mechanism for the following problem, and for evaluating the validity of some analysts' choice of a measure of effectiveness.

During World War II, the British armed their merchant ships in the Mediterranean Sea with anti-aircraft guns to fight off enemy aircraft. These guns were in short supply, ex-pensive, and badly needed elsewhere. After a few months of operation on the ships, the British government ordered an analysis to decide whether the guns should remain on the ships. Using a systems approach, the analysts' model looked like this:

After considering this information, the British government decided to remove the guns from the ships and redirect them to more gainful employment. Fortunately, before the decision was implemented, someone pointed out that the wrong measure of effectiveness was used to provide the feedback.



The objective was to protect the merchant ships, not to destroy enemy aircraft—that could be done more efficiently in other ways. The guns, however, forced the attacking aircraft to maneuver more, release their bombs at higher altitude, and otherwise impaired the bombers' accuracy. The guns were serving their purpose because more cargo was arriving. When the MOE (feedback) was framed correctly against the decision objective, the analysts discovered only ten percent of the gun-protected ships were sunk during air attacks while twenty-five percent of the unprotected ships were lost. Based on this revised analysis, the British left the guns on the ships.

8. Adapted from *Methods of Operations Research* by Philip M. Morse and George E. Kimball (Cambridge, Mass.: MIT Press, 1951).

RELIABILITY

Next, we evaluate our set of criteria for reliability. In our lexicon, reliability is the accuracy and consistency of a measure. *How well can we measure?* We must specify to the analysts the resolution of the measurements we require, including the units of measure and the fidelity or degree of accuracy we desire for each measurement. We must tell them how much measurement error is tolerable. When we measure repeatedly, under identical circumstances, we should get the same, consistent results.

We select criteria with less engineering precision (resolution) to support decisions by the Secretary of Defense than we would for an acquisition program manager. Similarly, we are usually less specific during concept studies and become more granular as we approach production parameters. Do we need to know airspeed in terms of Mach, knots, or feet per second? Is greater precision of value to the decision maker or is it a distraction? The resolution we need to distinguish between alternatives in a meaningful way is the level of detail we should measure and display; this may be considerably less than the resolution we can possibly measure.

Ideally, we opt for criteria that we can measure directly, in isolation, and without disruption by the act of measuring in order to minimize error and improve repeatability. Measurement error is ever present; we can compensate for some measurement errors easily, such as that in a gauge that misreads by 10 psi across its entire range. Detecting or adjusting for other measurement errors is difficult, especially as our criteria become interrelated or more subjective. A missile's failure to intercept a target within lethal range (miss distance) is a typical test criterion. Test firing intercept failures may be due to hardware casualties in a sensor in the missile seeker head, problems in the missile's software, or its control system; but we cannot isolate the fault unless we measure at each control point. In the worst case, errors may cancel each other out and our miss distance may be small enough to score as an intercept even though the missile did not work properly. Miss distance does tell us something we want to know; it has high validity. Miss distance, if we measure it simply as distance from the target, has low reliability because we do not know how subsystem measurement errors interacted with one another or how they individually affected overall system performance.

Surveys present our most difficult reliability challenge, a circumstance wherein reliability is on a par with validity. When we commission surveys of personnel to research policy options, the quality of the questionnaire is central to the reliability of the results, so we test the questionnaire before we use it in a survey. By issuing the questionnaire, then interviewing the respondents and identifying why they answered the way they did, we gain confidence that responses from the general population mean what we think they do. If the questions are poorly worded, the respondents' answers will be skewed, compromising reliability. Reliability also suffers when we do not get a sufficiently large or random sample of the target population; we should not permit self-selection by respondents because the most vocal members of the population are seldom the most representative of the general population. Reliability suffers further when survey respondents do not answer truthfully, i.e., without necessarily meaning to be deceitful, some people answer questions based on how they think they should feel rather than how they actually feel. An old saw says voters speak from the heart but vote from the pocketbook; a similar process can happen with answers to surveys.

Sometimes, when we are assessing complex or intuitive behavior, there are limits to the amount of knowledge we can obtain about causal factors or future actions. When we compare two manufacturers' products based upon their anticipated mean time between failures, we can examine historical data from the companies, we can review their assumptions for calculating projected failure rates in the past, but we cannot know if their estimated rates are correct (reliable) until the product is built and tested. Even then, we still have uncertainty. Will the mass-assembled products behave like a lab-built prototype, or even like each other? We will explore this further in Chapter 5, "Uncertainty and Risk."

We desire repeatability or consistency, the same results under the same circumstances, in our measurements of criteria. We may not be able to reproduce the same circumstances for each measurement, just as downhill skiers race on a slightly different course on each run. The more subjective our criterion is and the more dependent it is on the actions of others, the less repeatable it becomes. The mood of a respondent to a survey question may alter his choices on any given day. In a conflict simulation, the enemy response may vary depending on which analyst is playing Red, affecting Blue's optimal strategies and outcomes dramatically.

We can improve reliability in several ways. First, we can measure the same criterion in more than one way. If we decide unit manning is a criterion for selecting which of several like units to deploy, we can examine overall strength, manning levels for mid-grade Noncommissioned Officers and above (leaders), and projected rotations during the deployment. Together, they provide a better picture than any one measure alone, and they all concern manning. Should we make each a criterion by itself? We could; it depends on the situation and the level of detail the decision maker wants when we model this problem. More likely, we will measure these three items to justify our evaluation of manning and display only manning in our briefing; if asked, we are prepared to explain our evaluation—the proverbial back-up slides.

A related way to improve reliability involves taking advantage of surrogates that we chose in searching for valid criteria. To see this point, recall the example of using various physical attributes of nuclear forces as surrogates for those forces' deterrent effect. To the extent that we can measure those attributes objectively, we can improve reliability. (Of course, we can only increase reliability if the attributes we measure are also valid measures of what we care about.)

We can enhance reliability by improving our measurement methodology. Improved measuring equipment with more sensitive instruments, more complicated models, or a more isolated test environment will lead to more accurate measurements. If we are using computer simulations, we can run more iterations. If sampling is an important technique, then we increase the sample size.

PRACTICALITY

We evaluate our criteria from a third perspective, practicality. *Does the knowledge we gain from measuring justify the resources that we consume?* Practicality in this application does not mean "easily used or applied," rather, are our criteria too costly to measure and use? Resources can be money, time, personnel, equipment, and the like—anything we consume to measure a criterion. Practicality involves a sense of the first two evaluations: Do we have enough validity and reliability? Can we afford more?

For example, there have been an enormous set of attributes that helped us to compare between the two prototypes of Joint Strike Fighters proposed by the two competing contractors. After we order them in terms of validity, practicality tells us how many are enough. We may be able to measure each to an extraordinary degree, and thus improve reliability, if we are willing to consume a large amount of resources to do it. Practicality considerations tell us whether we should. An example of a low level of practicality is a set of criteria that is both highly valid and reliable, but that requires more time to collect the data than is permissible to meet the deadline for this analysis.

Practicality may involve a tradeoff between validity and reliability. We can often improve both validity and reliability by consuming more resources. To conserve resources, we can choose more abstract, less costly, surrogate measures as long as they have enough validity and reliability to support our decision. Practicality constrains our analysis by tying it to resource limitations commensurate with the importance of the decision to our organization.

VALIDITY, RELIABILITY, AND PRACTICALITY INTERACTIONS

Having discussed validity, reliability, and practicality at some length, we should reflect on how they interact and how they are distinct from one another for they are recurrent themes that permeate our decision-making framework. Logically, we view and evaluate them sequentially. Validity is often our first and most central concern. When we analyze a problem and its alternatives, we are analyzing an abstraction of the real world (a model) and validity is our evaluation of how well we have transferred reality to that model. Without valid criteria, there is little point in proceeding further; the most exquisite reliability cannot compensate for measuring the wrong things.

Reliability, then, is our next concern: poor reliability can lay to waste a perfectly valid model in several ways. If we measure poorly or inappropriately, our data is skewed and our analysis becomes tainted. Flaws in reliability may be more insidious than validity problems because they are not necessarily obvious when the results of analysis are documented and displayed. We must insist the analysts show us how they measured before we can have confidence in their results.

Practicality can be viewed as resource allocation between validity and reliability. Often, we would like to measure more criteria and often we would like to measure an individual criterion with more precision. Practicality is the balance between the two: are we measuring so many things that our reliability suffers too greatly from spreading ourselves too thinly? Are we omitting an important criterion because we are measuring the others in more detail than we need? Are there insufficient resources to support this analysis and bring it up to the standards we need to achieve acceptable validity and reliability? Most of our practicality problems can be resolved with more personnel, time, or money. Our practicality evaluation tells us whether such expenditures are worthwhile in the context of the decision and the organization.

Finally, validity, reliability, and practicality are not absolute qualities that are either present or absent; *criteria do not pass or fail a "Validity-Reliability-Practicality Test."* Simple statements that declare, "A criterion has high validity because it reflects the real world" are not helpful; we must consider all of a criterion's characteristics before we are satisfied. There is an important, deliberately subjective quality to our assessment of these traits—we evaluate validity, reliability, and practicality from our decision maker's perspective. Therefore, we are not surprised when other organizations and other decision makers select or emphasize different criteria. Because of practicality constraints, the decision maker must approve the decisions we make about criteria, including which imperfections in validity and reliability are tolerable.

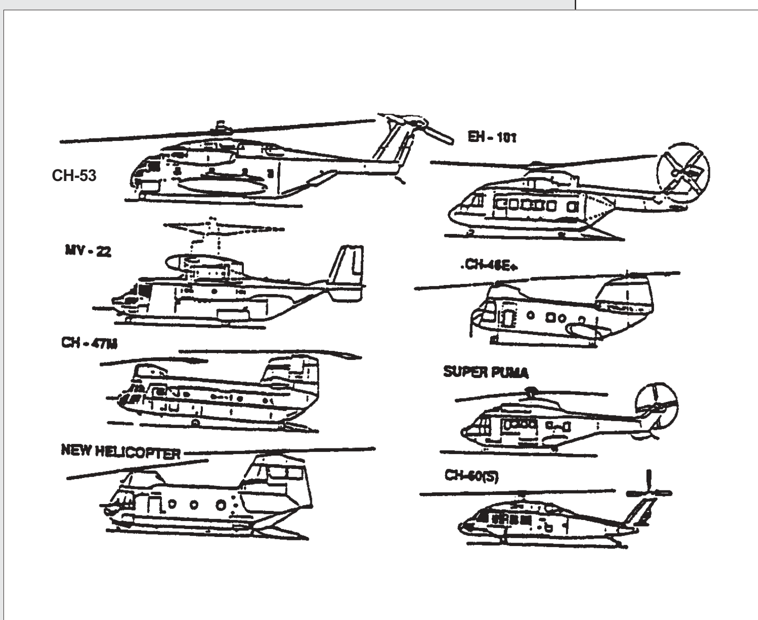
Measures of Effectiveness

We know that effectiveness is the ability to produce a result we desire, but there is usually no single measurement that will encompass all of the attributes we desire to measure in a set of alternatives. Speed and tire pressure are both attributes of an aircraft. One is clearly more important to the success of a fighter aircraft than the other. Speed contributes toward success in combat; it therefore becomes a criterion as a measure of effectiveness. Tire pressure is an attribute of the tire and ultimately of the plane, and it is a requirement for the proper functioning of an aircraft with inflatable tires. It is not a criterion that helps us evaluate how well an alternative satisfies the analytic objective. We do not care what the tire pressure is as long as it is adequate. It is possible to imagine a case in which every MOE of several aircraft is exactly the same, thus tire pressure emerges as the tiebreaker, but such circumstances are rare. (They might be more common in shopping for less expensive items. Color might be the discriminator among several suitcases that all have the same capacity). If such a condition occurs, we might ameliorate it by more accurate measurement of more important criteria.

Criteria for procurement decisions thus tend to cluster around MOEs such as speed, range, capacities, weapons loads, combat power, lethality, and survivability. Note that we can measure some of these by direct means; others may require sub-measures to evaluate them meaningfully. We can measure the speed of a vehicle directly. The survivability of an armored personnel carrier may require a compilation of other measures like thickness of armor plate, profile, self-defense capability, redundancy of systems, etc. Note again that "self-defense capability" may require further specificity, such as the performance characteristics of an offensive capability such as a machine gun.

CASE STUDY: THE ANALYSIS PHASE—ALTERNATIVES AND MEASURES OF EFFECTIVENESS USMC MEDIUM-LIFT REQUIREMENTS: THE V-22 OSPREY AND HELICOPTERS

Congress and the Department of Defense specified many of the aircraft options the Institute for Defense Analyses considered in their analysis of medium-lift alternatives, but they gave IDA license to explore other alternatives as well. Thus, Congress and DoD wanted IDA to consider the broadest range of options; they are shown above. As a result, IDA added the New Helicopter, a notional design based on a Boeing 360.



Alternatives

For each aircraft, IDA created two fleets for their assessment, one sized on the Marines' requirement to lift the assault force in two waves of aircraft (502 V-22s) and the other sized on the projected expenditure by DoD for replacement helicopters (356 V-22s). In each case, they calculated the cost of the V-22 fleet and used the same funding level to buy the various helicopter fleets. All of the fleets were viable in the sense that they were plausible alternatives, however by fixing cost at these two levels, IDA did not evaluate whether a helicopter fleet less costly than the DoD proposal in Level II could achieve the mission, i.e., they used DoD's planned expenditure as a lower boundary. No reasonable options were

9. Simmons, L.D. et al, *Assessment of Alternatives for the V-22 Assault Aircraft Program*, Executive Overview, Institute for Defense Analysis, 1991, p. 12.

excluded from the study; in fact IDA underplayed some significant additional costs to keep the smaller helicopters in play.

No matter which aircraft is selected for procurement, the Marines' existing fleet of 76 CH-53E heavy-lift helicopters must augment the medium-lift fleet. Some of the smaller helicopter fleets would require *additional* CH-53Es. The smaller helicopters cannot lift certain "medium" weight cargos such as vehicles and artillery. As a reference point, at the time IDA did their study the Marines had 224 CH-46E medium-lift helicopters and 76 CH-53E's. Table 4 below reflects the size of the fleets at the two cost levels IDA considered:[9]

Marine Corps Medium-Lift Assault Aircraft	Number at Cost Level I ($33B FY88)	Number at Cost Level II ($24B FY88)
V-22	502	356
New Helicopter	634	450
CH-47M	673	527
CH-60 (S)/CH-53E+	287/347	240/283
CH-46E+/CH-53E+	317/336	251/258
Puma/CH-53E+	330/322	260/246
EH-101/CH-53E+	252/335	200/256

MEASURES OF EFFECTIVENESS

Congress and DoD together identified eight missions that they tasked IDA to evaluate. IDA evaluated the role of the aircraft in each mission area and explored the comparative performance of each aircraft fleet using the following MOE:

- *Amphibious Assault* (Move Troops and Equipment Ashore). IDA's MOE was the percentage of the assault force lost while building a 3:1 force superiority during a vertical assault. They used survivability of the different aircraft in the assault role as a proxy. Using aircraft speed, design, and size, IDA evaluated how likely enemy air defenses were to shoot down the aircraft under a variety of conditions, e.g., day, night, rolling and flat terrain, various air defense weapons. The defending force was a Soviet-style, Third World Motorized Rifle Division.

- *Sustained Operations for Logistics Support* (Move Troops and Equipment to Support Combat Forces Ashore). IDA compared the number of equivalent payload sorties flown in a 30-day period, based on aircraft reliability rates, payload, and speed for the different fleets of aircraft.

- *Hostage Rescue/Raid* (Insert and Extract Marine Rescue or Raiding Force and Hostages). For this mission, IDA evaluated the maximum distance from the objective a raid could be launched and, separately, how long it would take to reach an objective from a distance of 275NM, the V-22's most distant possible launch position. The helicopters had to have their ships close toward the objective before they could launch.

- *Overseas Aircraft Deployment* (Move to Overseas Theater and Transport Deployed Marine Force to Combat Positions). IDA assessed the number of C-5 sorties required and how long it would take to deliver a brigade's share of each fleet to an off-loading Maritime Pre-Positioning

10. Since 1990, the Marines have adopted Operational Maneuver From The Sea as their operational concept and it calls for Over-The-Horizon amphibious assault, incorporating the V-22 to land the vertical assault echelon from up to 50 NM offshore.

Squadron or to the Marines' pre-positioned brigade equipment set in Norway. They also evaluated how long it would take the aircraft to deploy and tactically reposition combat troops.

- *Combat Search and Rescue* (Recover Downed Air Crews). IDA evaluated the percentage of rescues each type of aircraft could affect within two hours of a crash based on the distance of the survivors from the launch platform.

- *Special Operations* (Insert and Extract Special Operations Forces). Clandestine Special Operations often require aircraft to over-fly hostile territory at night, therefore IDA compared the fleets based on the number of missions that each could complete in darkness during nights of varying length.

- *Counter-Narcotics* (Trail Courier Aircraft and Boats, Deploy Law Enforcement Personnel). IDA evaluated the area to which each aircraft could respond in three hours and at maximum range without refueling.

- *Anti-Submarine Warfare* (Detect and Attack Enemy Submarines). IDA compared the V-22 fleet's capability using dipping sonar to detect submarines approaching the battle force to that of the Navy's S-3 patrol plane fleet (with other sensors).

See Appendix 3 for the results of IDA's analysis of each MOE.

Validity. IDA used a plethora of labels to measure the same thing in all eight missions: speed. This is a classic example of how seemingly different criteria can, in fact, be different representations of the same thing. Cycle time, area searched, time over an area, and the like are different measures of speed. This is why much of the IDA analysis seems repetitious.

Although we normally seek criteria that are unique, is the use of non-unique criteria justified in this case? Yes. Speed is a dominant criterion in each of the missions. The V-22 is more effective because it is faster; it is also more costly, as we shall see. Again, the crux of the decision is whether the additional effectiveness derived from the V-22's higher speed is worth its cost. With the IDA study, the validity question we should really ask is whether each scenario is truly representative of medium-lift aircraft employment: our standard question becomes, "Did we measure the right thing in the right context?"

The Marines validated the assault scenario, the most important medium-lift mission by far. It drives the overall size of the medium-lift fleet.[10] Survivability is an appropriate proxy for estimating how fast combat power will build up. Looking at how well each aircraft supports Over-The-Horizon assault was critical, and one could argue (despite the Congressional and DoD guidance), it is the only scenario that really merited evaluation. The sustainment scenario is based on how many sorties each aircraft can generate vice how many sorties and how much equipment the Marines require for support. This makes the measure of sortie rate questionable in terms of validity because the superior performance of the V-22 may not be necessary to achieve the mission, i.e., it may be over-capacity.

The Hostage/Raid scenario starts with the amphibious ships at the V-22 launch point and includes the steaming time for the ships to close launch points in the helicopter response times. To judge the validity of this MOE, one must examine the historical record for instances in which operations were delayed or canceled because of the additional ship transit time and then look at our current and projected needs. For example, the Marines have shown how the aborted Desert One raid and Non-Combatant Evacuations could have been executed more easily with V-22s. Our validity question is whether the 275 NM scenario, based on the operational range of the V-22 vice real world data bases and planning scenarios, will happen often enough in the future for it to be used as the test case in this study. If most operations will begin 1500 miles from the objective, the relative response time difference is much smaller between different kinds of aircraft. If the predominant circumstance is that the ships are already nearby, then again the response time difference between aircraft types is quite small. We can tell that IDA's chosen scenario favors the V-22, but we cannot tell with the information available whether that kind of scenario is itself sufficiently valid.

The self-deployment scenario shows a clear advantage to the V-22. Less need for high demand supporting strategic airlift is important—and the earlier arrival of the V-22 to move troops is markedly better than the helicopter options… provided the 250 C-141 sorties of the Fly-In Echelon of the Marine Expeditionary Brigade arrive in time for the Marines to be transported by the V-22s.

For the non-USMC missions, speed is still the dominant criterion IDA used to compare aircraft options. Where range is concerned, the V-22 flies further because it flies faster every hour it is in the air, a significant advantage over helicopters. For Combat Search and Rescue, speed is indeed of the essence and its validity is strong for estimating success. For the long-range Special Operations missions, IDA assumed the assault force started at a great distance from the objective, and they assumed that more Special Operations are better. But the V-22 may again represent excess capacity: are more Special Operations required and are planners limited by the current inventory of helicopters?

For counter-narcotics operations, the response times from cueing to aircraft arrival in order to trail boats and aircraft or to move agents to a site is a highly valid criterion for an individual mission, much like for Combat Search and Rescue. We must ask, however, whether there are circumstances under which it would be more advantageous to have two less capable aircraft rather than one V-22.

Submarine detection and localizing (vice area searched based on speed of the aircraft jumping between dip points) is the most valid way to compare anti-submarine warfare systems because it is the most difficult chore in the detect-to-engage sequence. All the aircraft alternatives carry similar sensors and weapons.

Reliability. IDA measured their MOE well, using existing data for aircraft characteristics where available and they scrutinized projected aircraft characteristics from contractors care-fully. IDA used military judgment from the Joint Staff and services to evaluate the subjective elements of the study such as the scenarios and missions, thereby improving the reliability of their analysts' estimates. The main reliability issues again revolve around the scenarios; did IDA measure aircraft performance accurately and consistently?

For the assault scenarios, IDA ran hundreds of iterations using the different fleets under varied simulated conditions to build a very large database. Field-testing the V-22 was not possible; however, the Marines had data based on helicopter-landed assault forces that IDA extrapolated to build the simulator runs. Scenario construction in terms of terrain, environmental conditions, and density of air defense along flight routes must all be realistic in order for the results of the simulation runs to be highly reliable; in this case they were as good as possible in 1990. The only way to improve reliability further in the all-important assault mission would have been for IDA to construct additional scenarios with a greater variety of opponents.

The outcome of the sustainment scenario depends upon the time between failures for the aircraft, i.e., how many round trips can each aircraft make with how much cargo before they go down for maintenance? The failure rates of the yet-to-be-built aircraft had to be estimated. IDA doubled the contractor's estimate, yet their calculations were still optimistic for a new (high risk) technology; in the IDA study, the V-22 was still more mechanically reliable than the advanced technology helicopters. What happens if the V-22 fails at triple the projected rate? How did the contractor estimate the failure rate in the first place? Historical research may reveal a trend between aircraft manufacturers' predicted failure rates and their actual failure rates. IDA could use such a factor as a better multiplier than simply doubling the contractor's estimate.

For the Raid/Hostage Rescue and Overseas Deployment missions, the reliability of the study is very high: we can predict the mission transit times and the aircraft are equally affected by environmental factors. The reliability of the measures for the Combat Search and Rescue and long-range Special Operations scenario is also high because it was calculated on the basis of the speed difference between the options, a straightforward mathematical process. For the counter-narcotics mission, the calculations of area coverage are similarly very reliable. For the Anti-Submarine Warfare mission, IDA's figures for detecting submarines are questionable because we do not have an explanation of how they calculated them.

Practicality. IDA took a very pragmatic approach to this study because they had to complete it quickly. They maximized their use of existing force-on-force models and data from previous studies and researched when they found them lacking. For example, earlier studies did not consider survivability in the assault scenario. They balanced knowledge gained versus resources consumed extremely well, achieving very high levels of practicality.

Validity and reliability for the assault scenario are in balance; improvements to either would be costly and time consuming beyond their worth. It is appropriate that it consumed the majority of IDA's resources; improving validity and reliability for the other scenarios by consuming more resources is not very worthwhile unless the scenario or MOE is grievously flawed.

For the sustainment scenario, estimating a better factor for anticipating failure rates (described above) to improve reliability, or doing sensitivity analysis using a variety of failure rates was probably worth the investment. The most important improvement to the Raid/Hostage Rescue and Overseas Deployment scenarios would have been for IDA to determine whether the scenarios they used are truly representative of how we anticipate these missions unfolding in the future. If most missions, raids, and rescues do not fit the IDA scenario profile, then we need a larger or different family of scenarios. IDA should have reviewed the theater Commanders'-In-Chief Operational Plans that will tell them quickly whether the Fly-In Echelons are expected early enough to take advantage of the V-22s' earlier arrival.

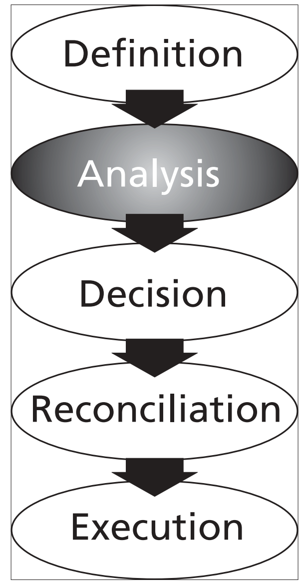
Summary

Executing the Analysis Phase forces us to answer some fundamental questions about how best to proceed: first about how much research we need to satisfy our analytic objectives and then what general approach we will take. For each analysis, we identify a likely range of viable alternatives that will reasonably satisfy our requirements. Sometimes we know them in advance and at others we decide upon them later in the Analysis phase, after we build our model.

We select criteria meaningful to the decision maker from the family of attributes that describe our options, beginning with Measures of Effectiveness. We evaluate each MOE for validity, reliability, and practicality individually and then the collection as a whole before moving on to address cost.

CHAPTER 4

ANALYSIS CONCEPTS: COST



What is a cynic? A man who knows the price of everything and the value of nothing.

-Oscar Wilde, *Lady Windemere's Fan,* 1892

COST IS ALMOST ALWAYS A CRITICAL FACTOR in defense decision making. Whether we are deciding on a new force mix, looking for a solution to a mission area deficiency, or choosing among policy options, someone during the decision making process will want to know the cost of our proposed solution. Selecting the best alternative and ultimately the success of our program or policy may well hinge on our ability to measure cost accurately and satisfactorily.

When we choose wisely, cost is on the opposite side of the coin from effectiveness. If we want to improve effectiveness, we will increase cost. If we cut cost, we reduce effectiveness. While we often discuss each separately, sometimes in isolation, they are inextricably related. The tension between cost and effectiveness is one of the reasons our defense resource allocation process is based on advocacy and adjudication: we fully expect the operators to demand the highest levels of effectiveness while the managers try to spread limited resources among a mix of programs to provide the best overall capability to all the operators, present and future. Indeed, this is the crux of the argument between supporters of the V-22 and those who wish to cancel it: is the greater effectiveness worth the additional cost—including the lost opportunity to fund other programs? (See Dr. Chu's testimony in Appendix 3.)

Selecting Measures of Cost

Measures of cost are a subset of the criteria we use to compare alternatives. Just as with measures of effectiveness (and all other criteria), we should have the decision maker approve our MOCs before we begin the analysis. There are two guidelines we follow when selecting measures of cost in addition to those for other criteria:

- Future Costs
- Standard Metrics

In addition to the immediate costs of alternatives, the cost portion of the analysis should also focus on costs yet to be born. We should isolate near-term costs and display them for the next budget year and, with only slightly less precision, up to the end of the next Future-Years

Defense Program (five or six years distant). In almost all cases, our analysis should consider life cycle cost in constant dollars.

We may include costs not measured in dollars, such as manpower, material resources, etc., as the situation warrants. We should use the same yardstick for each of the alternatives to permit easy, side-by-side comparisons. This means we must specify the type of cost information we require when we issue a request for proposals; respondents naturally tend to emphasize the type of cost most favorable to themselves and, left to their own devices, they may not provide the same types of costs as one another. We test our selection of MOCs, individually and as a set, using the concepts of validity, reliability, and practicality that we discussed in Chapter 3.

For an executive decision maker in the Department of Defense, the issue is not the mechanics of calculating costs; the point is whether the measures of cost proposed by the analyst fit the problem. Senior leaders in DoD must also be prepared to provide guidance to the analyst with respect to how they want costs estimated, lest the cost information they receive not support rational decision making.

Types of Cost

Cost is what we give up for what we want; our opportunities forgone. Money is the most common way to measure costs, but other methods are often more appropriate for force planning decisions. Ships burn fuel, expend ordnance, and need people to man them. Analysts can turn all these into dollars, but in combat the resources themselves are more direct and appropriate measures of cost. Many times we need to recognize that these other kinds of costs factor into peacetime force planning as well. The space a ship takes up alongside a pier or the wear and tear on an aircraft-launching catapult are costs that dollars alone cannot describe accurately. Cost, in addition to resources consumed, also represents opportunities lost by the choice of this use of money.

Analysts add modifiers to specify a multitude of specialized types of cost. The Navy's *Economic Analysis Handbook* alone has three pages of cost definitions. As executive decision makers, we need to understand the fundamentals of cost terminology in order to compare alternatives accurately and to communicate clearly among ourselves and with analysts. Most of our discussion of types of costs concerns procurement options, but many of these same concepts apply to policy alternatives.

We must be certain that the contractors, the analysts, and we ourselves use the same terms and define cost the same way. During a competition among American companies for a recent foreign military sale, the purchasing nation wanted a tactical aircraft that could deliver infrared-guided, air-to-ground weapons. One aircraft had this guidance capability built into the fuselage and nose of the aircraft. The other aircraft used a detachable pod carried under the wing. Both manufacturers' aircraft therefore met the requirement. The decision makers sought to compare the two alternatives' flyaway costs. The latter aircraft's manufacturer did not include the price of the pod in its proposal because of the loose way the purchaser defined cost, significantly (and knowingly) decreasing the apparent cost of that aircraft.

RELEVANT COST

One of the most difficult and important concepts of cost is differentiating between costs that result from a decision and those that do not. Relevant costs, as we define them for this course, are *forthcoming* costs that distinguish among the alternatives in our decision. They include the costs common to all the alternatives and the unique costs of each. While all costs are relevant to one decision or another, we tend to focus on costs that concern our organization. The concept of relevant costs is akin to that of validity—our need to ensure we measure the things that matter.

For example, the Navy programmers who estimate the cost of a new aircraft carrier usually exclude the cost of the air wing and surface ship escorts from the cost of the ship, even though the carrier cannot operate effectively without either. The cost of the aircraft and escorts are irrelevant to the cost of the aircraft carrier—so long as they do not increase or decrease because of our decision. If, however, we pick a new aircraft carrier design that requires five more aircraft than another design, the extra cost of the five aircraft is a result of our decision, and thus relevant. Before discarding any cost as irrelevant, we must be absolutely certain that our decision is not concerned with it.

SUNK COST

Irrecoverable expenditures we have already made are Sunk Costs. They are irrelevant to our decision because we cannot recoup them no matter how hard we try. Sunk costs are useful (in a historical sense) to determine the actual cost of an activity or program and help us predict the overall cost of new proposals. They are also very important for legal and accounting purposes, but not for decision making per se.

Sunk costs, in and of themselves, should have no bearing on an economic analysis or a decision concerning the future expenditure of resources. One problem with sunk costs is that we are not perfect economic people. We have a natural tendency to see value in money already spent, and, especially, in our effort already expended. We do not want to consider the time we devoted to a project as wasted, so we are inclined to continue programs and policies after we have spent money on them, even when the current course of action is no longer the best alternative. This situation occurs most often when the original need for a program has diminished or disappeared. For example, the decline of the Russian Navy as a blue water competitor with the U.S. Navy has led to major changes in U.S. naval strategy and doctrine and thereby the restructuring of many programs. Some programs have slowed down and others have been canceled despite the resistance of their well-intentioned program managers and community sponsors. Executive decision makers made their force planning choices about which programs to continue based on the requirements, urgency, and future costs of the alternatives. Pleas based solely on sunk cost rightly fell on deaf ears.

Programs that become more advanced, however, often gain a tangible advantage over competing alternatives as their sunk cost accrues. Since we focus on future cost, a new alternative, with all its research and development costs before it, is unlikely to be competitive from a cost standpoint against a program that is already underway. That is why the most acrimonious debates in defense resource allocation occur when we decide which programs to start.

GETTING THE CAMEL'S NOSE UNDER THE TENT: THE F-22 RAPTOR

Advocates know that the money spent before a system is operational generates program inertia, which makes a funded alternative increasingly preferable. As more of its cost becomes irrecoverable, other options become less competitive on the basis of cost alone. A classic example of this is the ongoing debate over the Air Force's acquisition of Lockheed-Martin's F-22 Raptor.

There has been nearly universal agreement since 1985 that the Air Force must re-place its fleet of F-15 air superiority fighters because of their advancing age. In the early 1990s, opponents of the F-22 argued that its cost per aircraft ($160-180M) was prohibitively expensive and deliberately underestimated by its proponents at $85M. The contrarians preferred continuing the production line of the older F-15 at about $45M per aircraft and cited a number of studies that the F-22, in conjunction with other tactical aviation plans, was unaffordable. The Air Force stressed the (undisputed) greater capabilities of the F-22 and determined they would find a way to afford it.

The Air Force focused largely on lesser near-term costs and built program inertia, despite Congressional concerns about cost. In 1997, Congress placed a $37.9B cost cap for procuring 339 fighters. That cap caused several major adjustments to the Acquisition Program Baseline, including December 2000's reduction from 86 to 73 aircraft in Low Rate Initial Production (LRIP) in hopes that full production aircraft would be significantly less costly.[1] In 1999 Chairman Jerry Lewis led the House Appropriations Committee to remove procurement funds for the F-22 from the FY00 Defense Appropriations Bill. The compromise that restored the funding required the F-22 to pass certain exit test criteria by 21 December 2000; however, Lockheed-Martin did not complete the required avionics and fatigue testing portions until 5 February 2001, in part due to bad weather and a labor strike.

In spite of this, Congress allowed the Pentagon to release $350M to keep suppliers and preparations for production going through 31 March 2001. The F-22 passed the Defense Acquisition Board's Milestone III on 6 February 2001. On 15 August 2001, the Defense Acquisition Board authorized LRIP of ten F-22s. However, its increasing cost mandated that the total number of airframes would shrink from 339 to 295. Based on usage of the term *cost*, both sides were vindicated: the *average* cost is now up to $173 million. However, the *variable* cost of the fighter is $84 million. The average cost, of course, included sunk costs of upwards to $20 billion in research, development, test, and evaluation funds.

OPPORTUNITY COST

Limited resources create opportunity costs; they are the things we forgo by choosing to attain something else. With unlimited resources, we have no opportunity costs because we can obtain all we desire; for DoD this would mean an unbounded force structure with all our programs and policies funded at 100 percent. In a world of constrained resources we must make choices: we fund more modernization than infrastructure; we deploy forces here instead of there; we fund recruiting incentives at the expense of retention bonuses; etc.

Decision makers in business measure opportunity cost most often in dollars, as profit made or lost. Imagine there is a factory that currently makes a product that generates profits of $100,000 every year. The corporate owners are considering retooling this factory to make a new product. The opportunity cost of surrendering the first use of the factory to make the current product should be added to all the other costs of beginning to make the new product; that $100,000 is forgone revenue and would otherwise have been added to the firm's profits. It is a cost as legitimate as all others are. No accountant records it because it is an event that did not happen, but executives must know and consider it.

In DoD, some opportunity costs are difficult to express in dollars, but we consider opportunity cost in every decision involving spending. If the Marine Corps decides to spend one million

1. Tony Capaccio, "U.S. Air Force To Delay Some F-22 Buys To Control Rising Costs," Bloomberg.com, 18 Dec., 2000

dollars for new trucks, and the next best use for those funds is new mobile field kitchens and power generators for field headquarters, we assume some analysis occurred to pick one option over the other. Without profit as a measure of return on investment, what measure did the Marine Corps apply? Almost certainly, something less tangible called value or utility, and the decision was made based on the most benefit to the service goal of winning battles. Programmers face these decisions daily; they know the operators want the best system possible but that the increased effectiveness must come at the cost—and therefore effectiveness (or even existence) of other programs.

EXTERNAL COST

Costs beyond the problem's boundaries are external to the analysis. Because they occur outside our organization they are usually irrelevant to our decision. In many cases, excluding a particular cost means that the cost, in reality, is now included in another organization's budget—we effectively transfer it outside our organization. The other organization may be a private company, another group within DoD but outside our chain of command, or another branch of government. If we do not transfer the cost very far, it may reappear later in the decision, when our organizations fall under a common superior. Our decision to exclude and thus transfer cost is a spillover effect onto those other organizations and they may object to the transfer when we reconcile the decision. For example, when the Public Works Department of a facility increases its utility surcharges to tenant commands to cover its increased costs, it is transferring the costs to the tenants who are usually outside its chain of command.

We should indicate to the decision maker where we assume or impose cost transfers. For example, when the Joint Staff and a unified commander consider the cost of a security assistance program for another nation, they consider the cost to DoD. Many other costs for the program will be borne by the Department of State, but those costs are not relevant to our internal DoD decisions. We may encounter the effects of the costs imposed on the State Department later as we reconcile our proposal, but we do not use them to consider the alternatives in our decision.

FIXED, VARIABLE, AND AVERAGE COSTS

Fixed costs are expenses that we incur whenever we initiate a course of action. They occur regardless of the intensity of the action or the number of items we procure; for example, research and development costs are fixed costs. Variable costs change depending on how we execute our program, particularly as we alter total purchase quantity or annual production rate. Fixed costs are tied to factors unlikely to change, such as the size and cost of the daily operation of the production facility. Variable costs change conditionally, as with adjustments to the size of the work force or the price of materials.

Period costs are fixed costs that accumulate over time, regardless of the amount of product or service purchased. They are primarily wages and facilities-related costs that may conceal inefficiencies we can eliminate by adjusting the production rate, thereby reducing the total cost. For example, we may have a labor force that is working below its capacity that cannot be reduced because of the distribution of skills required to produce each item. However, if we have funds to buy more materials and we can accept earlier deliveries (which may mean training more DoD operators in the near term—spillover costs), the contractor could produce more systems in the time that the labor force is being paid. We reduce the total time to produce all the systems and save period costs by having fewer periods.

$$\text{AVERAGE COST} = \frac{\text{TOTAL COST}}{\text{TOTAL QUANTITY}}$$

TOTAL COST = FIXED COST + VARIABLE COST

IF THE NAVY BUYS…

12 PATROL CRAFT		6 PATROL CRAFT	
FIXED COST		*FIXED COST*	
R&D	$1.5M	R&D	$1.5M
FACILITIES	$12M	FACILITIES	$12M
ANNUAL OVERHEAD	$2.5M	ANNUAL OVERHEAD	$2.5M
	$16M		$16M
VARIABLE COST		*VARIABLE COST*	
MATERIALS	$24M	MATERIALS	$12M
LABOR	$16M	LABOR	$8M
	$40M		$20M
TOTAL COST	$56M	*TOTAL COST*	$36M
AVERAGE COST	$4.7M/CRAFT	*AVERAGE COST*	$6.0M/CRAFT

Figure 4-1. Fixed, Variable, and Average Costs of Patrol Craft.

Variable costs change with the ebb and flow of the production quantities and scheduled deliveries. A particular type of variable cost is Incremental Cost: the added cost of purchasing one more of something, e.g., adding one more destroyer to a program or one more student to each seminar. The fixed costs remain the same, but the additional unit requires more resources: more labor and materials for the ship and more administration, counseling, and grading for the student.

Generally, variable costs decrease per unit as purchase quantity increases, up to some threshold, e.g., the maximum fabrication capacity of a facility. These result in volume discounts where the manufacturer lowers his price per unit to reflect the wider distribution of fixed costs among more units. Prices may jump upward again if the contractor opens a new production facility (incurring new fixed costs) and begin declining again as production increases.

Average cost is the total of fixed and variable life cycle costs divided by the number of units we procure.[2] Let us suppose the Navy plans to buy new patrol craft for its units assigned to the U.S. Special Operations Command. Regardless of how many craft the Navy purchases, there will be unchanging fixed costs associated with the program: the design work, setting up an assembly area, signing and managing contracts, etc. These are summarized in figure 4-1 as fixed costs. Because of these fixed costs, reducing the production quantity from 12 to six boats, as shown in figure 4-1, increases average cost. The total program cost is reduced, saving money, but the average price per boat grows higher.

Why is this important? With the high cost of modern weapons systems, many critics of the Pentagon cite average prices in their argument. If we cut a program and reduce its purchase quantity, we will not recoup as savings the average price of the equipment forgone, just as dropping from 12 to six patrol craft did not halve the cost of the program in figure 4-1.

	FY 02	FY 03	FY 04	FY 05	FY 06	Total Cost	Cost per Craft
3 Craft per year	$13.5M	$12.5M	$12.5M	$12.5M	$12.5M	$63.5M	$5.3M
6 Craft per year	$13.5M	$22.5M	$22.5M	0	0	$58.5M	$4.9M

Table 4-1. Stretching Out Procurement.

Reducing the production rate, thus stretching out a procurement program over more time, is a common technique to reduce near-term cost. However, the overall effect of stretching a program is to increase the total cost of the program and the average cost of each system because we

2. Related to average cost, we often see aircraft described in terms of Fly-Away Costs and Procurement Costs. Fly-away costs include only research and development and narrowly defined production costs. Procurement costs include fly-away costs plus initial contractor support of production models, contractor training of service personnel, and an initial set of spare parts for each system.

must bear the fixed costs longer. Returning to the patrol craft example, instead of buying them all in one year, as in figure 4-1, we will consider two cost streams for purchasing 12 patrol craft. The first uses a production rate of three boats per year for four years and the second six boats per year for two years. Table 4-1 shows the start-up costs (research and development and facilities set-up) occurring in FY02 with production beginning in FY 03. The costs shown in FY03 through FY06 are the annual overhead of the boatyard plus labor and materials for each boat. By stretching out the boat fabrication over four years, instead of two, we increase the total cost of the 12-boat program by $5M or 8.5 percent.

Industrial fabrication has a phenomenon called the Rate Effect. It describes the way costs change as the production rate shifts away from full capacity and explains why cost reductions do not decrease linearly with cuts in the production rate. As we discussed earlier, only the variable costs are eliminated while the fixed costs remain. If we order fewer missiles than the contractor had anticipated, the contractor will not be able to proportionately reduce all costs and DoD will pay more per unit than we anticipated. This is why stretching out programs to reduce near-term cost not only increases average cost and frustrates planners, it also creates instantaneous procurement and life cycle cost overruns.

Why would we ever stretch out a program? Sometimes the pressure on the near-term budget is so great that we must reduce production rate to keep the program alive; the only other choices are to cancel it or another vital program outright because the money simply is not available for full production. The closer we get to the budget year, the more "real" the money becomes and the more necessary it is that we refine spending forecasts and push spending further into the future to balance the books in the near-term.

The cumulative effect of stretching out programs, as DoD has done over the past ten years, is extremely deleterious. Because fewer replacement vehicles and systems reach the operating forces, the average age of equipment increases. Older equipment requires more maintenance to stay ready, drawing resources away from modernization (and other) accounts. The effect is cumulative, too. As we delay purchases year after year, the total number of new procurements we need increases; the new deferrals add to the old, and now we face a department-wide procurement bow wave that analysts estimate will cost an additional $80-120B per year to maintain DoD's current force structure and replacements for aging weapons.

In complex decisions, the determination of which costs are fixed, which are variable, and the correct construction of average cost are critical to thorough economic analysis. Separating fixed and variable cost is very important when we make decisions about incremental changes to programs or policies. Just as DoD cannot save the average cost per mile by steaming a ship one less mile (we save only a part of variable cost by conserving the fuel), we cannot save the average cost of educating a Midshipman by decreasing the Naval Academy's enrollment by one.

LIFE CYCLE COST

As shown in figure 4-2, life cycle cost includes all the costs associated with a system from conception to disposal or deactivation. Note that the segments in the figure are additive (this is often called a sand chart) and that the top-most boundary is the combined cost for that time period. Many executive decision makers focus on procurement cost because they assume that it represents the biggest share of life cycle cost. Historically, however, the largest part of life cycle cost is for operations and maintenance during the service life of the equipment. For example, 80 percent of the life cycle cost of an average Navy ship goes toward operations and maintenance after it joins the fleet (and 50 percent of this cost is, in turn, for personnel).[3] For almost all procurement option comparisons, life cycle costs are among the most important criteria for defense decision makers. The major components of DoD life cycle cost are:

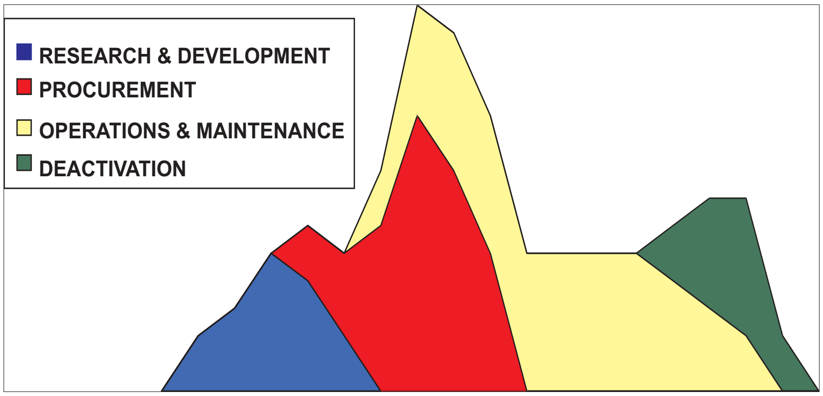


Figure 4-2. Life Cycle Cost.

- ***Research and Development Costs*** (3600 money[4]): concept and feasibility studies; engineering design; test, and evaluation of engineering models; and associated management functions.
- ***Procurement Costs*** (3080 money): industrial engineering, facility construction, process development, materials, manufacturing, production operations, quality control, and initial logistics support requirements.
- ***Operation and Maintenance Costs*** (3400 money): training DoD personnel; consumable supplies such as fuel, spare parts and other sustaining logistic support; intermediate and advanced maintenance, and replacements distribution.
- ***Deactivation Costs***: demilitarization, disposal of non-repairable items, system retirement, material recycling, and related logistic support requirements.

TOTAL OWNERSHIP COST

Beyond life cycle cost, DoD has embarked on a new attempt to capture during acquisition planning all the costs associated with hardware, i.e., the transfer costs borne by the users and owners of the equipment procured by the acquisition system. The DoD definition of Total Ownership Cost is: *Costs to research, develop, acquire, own, operate, and dispose of weapon and support systems; other equipment and real property; the costs to recruit, train, retain, separate, and otherwise support military and civilian personnel; and all other [related] costs of business operations of the DoD.*

Total ownership cost includes all aspects of life cycle cost and more; in addition to direct personnel-related costs (crews and their training), it includes the cost of the supporting infrastructure that plans, manages, and executes the program over its full life, as well as the cost of common support items and systems that a service incurs because of the introduction of the system. The Navy is using this methodology with 20 test programs in place, including some of our largest weapons systems acquisition programs. By exhaustively including second order costs and beyond, the Department of Defense is acknowledging that the greatest costs associated with many programs occur after the system becomes operational, that those costs should be considered when choosing among alternatives, and that therefore we need to find a way to capture them in advance to support analysis.

This means many costs we previously counted as external to a program are now internal, e.g., the educational and recruiting costs of the share of boot-camp recruits who are destined to work on a particular system. Program managers reduce total ownership cost through their tra-

3. J. Talbot Manvel, Jr. , "The Next-Generation Aircraft Carrier," *United States Naval Institute Proceedings*, Jun. 2000: 70.
4. Programmers in the Pentagon use these shorthand codes to refer back to budget rules that restrict how funds may be expended, e.g., “We will not be able to obligate all our 3600 money for this program by the end of the fiscal year.”

ditional attempts to reduce life cycle cost for their systems and now by reducing demands on the rest of the Navy. Thus, manning reductions in the crew and support staff of maintainers and logisticians has become a priority for the program manager whereas previously he or she focused almost exclusively on the cost of the hardware and perhaps spare parts. The program manager may decide to incorporate more expensive—in terms of procurement dollars—labor saving devices to reduce crew-manning requirements to reduce total ownership cost; under traditional acquisition management philosophy, he or she would be tempted to opt for the less expensive equipment to keep the procurement cost lower. Similarly, training commands that educate technicians are examining their courses to reduce the time to get the sailor to the job, lowering the average "cost" of a sailor and thereby the total ownership cost for the system.

Implementing total ownership cost concepts will not be easy. The Air Force owns and manages all of DoD's space assets, but all of the services use them. Should the Air Force charge user fees to the other services, similar to Working Capital Fund arrangements on bases, to reduce its total ownership cost? Because total ownership cost includes "linked-indirect" costs, i.e., those that are generated as a result of introducing and supporting a system, but which cannot be directly associated with one specific usage or program, where do we draw the line and prevent system-owning commands from charging expenses to a user that the owning command would incur anyway? Do we assess the average cost or the incremental cost of the manpower associated with the support system in the user system's total ownership cost? Should we allow the Air Force to include part of the cost of manpower recruiting, basic training of the technicians (and the recruiters), and electronics training common to all space systems in its user fees? Clearly, we would exclude the costs of Air Force marching bands and fighter squadrons, but there are many gray areas in between.

Three notions appear from the idea of total ownership cost. First, as we discussed above, although the concept of total ownership cost is clear, calculating, measuring, and centralizing these costs is difficult, raising a large practicality issue. Second, our economic analysis, particularly analyses of alternatives, could be hampered by simplified or uneven total ownership cost efforts. A single-seat attack aircraft halves the personnel costs of air-crews compared to a two-seat version of that system. Will cost predominate in this case at the expense of effectiveness? Third, many systems are themselves largely dependent upon other systems, or would not even exist without them. For example, how should we assess the total ownership cost of the Joint Stand-Off Weapon, an air-delivered Global Positioning System (GPS)-guided weapon? This weapon could not function without GPS; therefore its total ownership cost should include part of the cost of GPS. On the other hand, we would have established GPS regardless of the standoff weapon, so why should it be taxed for something that would have happened anyway? What is the fair apportionment of total ownership cost for each user of GPS?

Cost, Effectiveness, and Schedule

Cost, effectiveness, and schedule are familiar criteria in defense decisions. While time can be thought of as a cost, we can also think of it as a performance factor: we would almost always rather have a capability sooner than later. Whether we treat schedules as a subset of cost or effectiveness, or as their own criteria, depends upon the decision. When we construct a schedule of the outlays for a program—a cash flow—we are combining time and money. When we construct deadlines for achieving a level of performance, we are combining time and effectiveness.

How quickly we can execute our decision is directly linked to when money is available; funding is a prerequisite for executing the schedule. Sometimes, if we obligate more money faster, we can accelerate research and development, the procurement rate, or the date of initial operational capability. Technology may also constrain scheduling, as is happening now with national missile defense; we have money available but cannot spend it wisely until we overcome several technical hurdles.

Types of Dollars

To help us evaluate alternatives, the analyst may include costs based on several different kinds of dollars. The types of dollars we primarily use in defense decision making are current dollars and constant dollars. We show them in figure 4-3 and will explain how we convert between them.

CURRENT DOLLARS

We spend current dollars. As we budget for the future, we express our planned spending in the dollars of the year when we intend to make the outlay. We also know that most costs rise over time due to inflation, which reduces the value of today's dollar. The actual purchase price we will pay in the future (the amount we will write the check for), increased from today's price because of inflation, is measured in current or then-year dollars (they will be current dollars *then*). Thus, if an item costs \$100 in 2002 current dollars and there is a ten percent annual increase in prices (inflation), we will pay \$110 one-year later using year 2003 current (then-year) dollars.

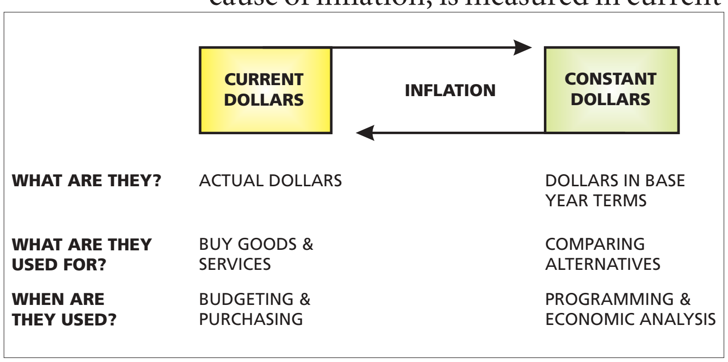


Figure 4-3. Types of Dollars.

All our authorizations and appropriations from Congress, including the Defense Authorization Bill and Defense Appropriation Bills, are expressed in current dollars because they represent the actual money we will spend. Likewise, the dollar amounts in the Executive Branch's federal budget are expressed in current dollars, as are those in the Future-Years Defense Program. Again, current dollars are the only dollars that are actually spent for goods and services.

Imagine that during the 2002 Defense budget preparation the Navy will request two identical ships, one to be built in 2002 and the other in 2005. Using a five percent price inflation rate,[5] a new ship projected to cost \$850 million in 2002 (2002 current dollars) would cost \$984 million in 2005 (2005 current or then-year dollars). Since the ships are identical, the increased cost is due to the rise in prices for goods and services alone: the effects of inflation from 2002 to 2005. We can use current dollars to compare values easily within the same fiscal year, but, because of inflation, current dollars are not useful for directly comparing and evaluating alternatives in different years. Is the 2005 ship worth more than the 2002 ship because it cost more? Obviously not, so we need a methodology to account for inflation so that we can examine the cost of alternatives across different years.

CONSTANT DOLLARS

There are many occasions when we wish to compare the price of equipment and services bought during different years. Inflation makes impossible an accurate comparison of worth, based on

5. DoD uses inflation rates provided by the President's Office of Management and Budget. If Congress disagrees, they may use their own inflation rate from the Congressional Budget Office and re-calculate DoD's math if they decide to authorize and appropriate the ships.

current dollars alone, because we must account for the inevitable erosion of purchasing power. Constant dollars are funds from different years that have been adjusted for the effects of inflation and benchmarked to a base year.[6] The base year may be any year we prefer—often it is the next fiscal year—the fundamental requirement is that we use the same base year for all our calculations. We can convert then-year dollars from a base year in either direction, forecasting costs into the future or reflecting into the past to make direct comparisons.

For example, suppose we desire to compare the price of a ship the Navy purchased in 2002 for $900 million (2002 current dollars) to an identical ship purchased in 1996 for $650 million (1996 current dollars) to see if there was a price increase beyond inflation. First, we determine there was a constant five percent inflation rate between 1996 and 2002, then we calculate the cost of the 2002 ship as if we had bought in 1996, our base year. The year 2002 ship cost $672 million in 1996 constant dollars. By converting costs to constant dollars, the analyst can say the true increase in price for the new ship is $22 million in 1996 constant dollars, not the apparent $250 million. Constant dollars provide a common measure we can use to compare alternatives independent of inflation so long as we calculate the costs of the alternatives using the same base year.

CASH FLOWS

When DoD buys a new weapon system or implements a new policy, we usually incur costs and make outlays over a multi-year period. We display our anticipated annual outlays in a table with years and budget amounts to create a picture of the program or policy's cost stream or cash flow. Cash flows facilitate comparisons among alternatives and are crucial as we prepare programs and budgets in the formal resources allocation process. Once we select a program alternative, its cash flow gets translated into budget lines as part of the Acquisition Program Baseline. Policy alternatives get funded similarly, after their cash flows are converted into the various types of funds Congress appropriates and DoD disburses.

Cash flows are different from life cycle cost because they describe only relevant (forthcoming) costs. For example, the Joint Strike Fighter Program Office authorized two consortia to produce flying prototype aircraft to compete for the production contract. The program office will compare the aircraft on the basis of cost and effectiveness; for the cost analysis, they will undoubtedly display cost as cash flows: how much it will cost to complete the program each year for the production run and service life of each alternative. They will also likely request cash flows for different production rates and total purchases.

Historically, life cycle cost or average cost may be of interest to some decision participants, but many of the costs of both prototype Joint Strike Fighter aircraft, such as research and development, are already sunk, and therefore irrecoverable. Based on our earlier discussion, they should not play into DoD's decision on how to proceed. Executive decision makers need to focus their attention on cash flows of relevant costs, those they will have to budget in the future. As sensible as this seems, for many decisions the senior executive in DoD must specifically request information on cash flows. Many analysts use older techniques and contractors prefer to present their options in the most favorable light—which may *not* be displayed as a cash flow.

---

6. The calculation is the same as for compounding interest: Constant Year $ = Base Year $ * (1 + i)n where i is the inflation rate and n is the number of years distant from the base year.

As we begin analysis of a new weapon system, we should have the analyst or contractor estimate the cash flow as a function of the production schedule and purchase quantity. Contractor costs are extremely sensitive to production rate adjustments, especially for major systems like ships and aircraft. Generally, as production rate decreases and delivery is delayed, fixed costs per unit rise. If the total procurement quantity is reduced, average unit costs again rise. Because we know that deviation from either the total number purchased or the planned production rate causes changes in cash flow, we may ask contractors for estimated cash flows for several purchasing strategies.

Contracting Strategies

Congress and the Department of Defense have long sought to minimize and prevent cost increases in procurement programs. There is a fundamental tension at work in defense acquisition. Contractors value stability—unchanging requirements (effectiveness), guaranteed production rates or purchase quantities, and predictable cash flows (outlays). The Federal Government, however, wants flexibility to modify a program—to improve effectiveness (often as the result of new technology) to give operators the best possible equipment and to change production rate or quantity to save near-term budget dollars. In business, one must pay for that flexibility by paying more money to the contractor in profits to reward his (and his shareholders) assumption of higher risk under more volatile circumstances.

The most common type of contract for new major weapons systems is "cost-plus" wherein the manufacturer bills the government for the actual cost of work done, plus a percentage of cost or fixed fee that is his profit. The need for oversight is obvious; the contractor has no incentive to reduce costs and in some cases incentives to increase them or at least charge as much shared corporation overhead as possible against a cost-plus contract. To provide an incentive for the contractor to reduce cost, DoD has begun to include performance incentives in its cost-plus contracts. The manufacturer receives bonuses for reaching program milestones ahead of schedule or for reducing costs below programmed levels. To increase oversight of these kinds of programs, each service has created corps of acquisition professionals who, when they are not assigned to their branch or warfare community, specialize in acquisition management. The DoD Inspector General, the General Accounting Office, and the Congressional Budget Office also exist, at least in part, to oversee government contract execution because of our stewardship concerns over acquisition programs.

Historically, DoD has made several major attempts to implement cost-saving strategies. In the 1970s, the Navy issued several "design-to-price" contracts for major weapon systems, including the Perry-class frigate. The key concept was to cap the production cost of each ship and thereby limit cost growth by making it fiscally impossible to add new equipment without removing something else; the new gear would have to compete for dollars and space within the existing design. The result was a ship that saw its capabilities progressively reduced, eaten away by inflation. For example, the fire control radars had less range than other missile and gun-equipped escorts because of their low power output. The fire control system was an off-the-shelf Dutch system, so the U.S. Navy had little flexibility to change its characteristics and even if it could, another capability would be lost to compensate. Predictably, the Surface Warfare community was dissatisfied with the result. Eventually, the ships were upgraded in a series of costly overhauls and design changes and they became effective escorts.[7] Operators' reactions to other design-to-price procurements were similar and the practice is currently disused—except, in a sense, when Congress places budget caps on programs like the F-22.

The second cost control method the government has attempted is to forgo flexibility by issuing fixed price contracts. In this case, DoD agrees to lock in requirements (specifications) and quantities that enable the manufacturer to predict his incoming cash flow and production requirements with near certainty. DoD uses fixed price contacts widely with bulk purchases and for simpler equipment. A government contracting office issues a request for proposals with quantities and specifications and the lowest qualified bidder wins the contract.

Secretary of the Navy John Lehman sought to extend this type of contract to major weapon systems in the 1990s. This, too, had unforeseen consequences. For major weapon systems, like nuclear-powered submarines, the specifications were written loosely enough to allow for the injection of new technology without renegotiating the contract. High-profile court cases resulted wherein the contractors alleged the Navy effectively changed requirements while refusing to compensate the contractor for the new costs thus incurred. As a result, many manufacturers will only accept fixed price contracts for mature programs, e.g., buying additional ships and aircraft that are already in production.

An important way DoD reduces the cost of a procurement option is to enter into a multi-year contract with the manufacturer that effectively locks in our cash flow for the period of the contract. Normally, DoD signs one-year contracts with a company to purchase a fixed quantity of goods or services. In some cases, we award the contract to different companies each year after competitive bidding. This process creates significant risk and uncertainty for the manufacturers as they try to predict their future sales and production level requirements. Multi-year contracts provide stability for the manufacturer by easing his uncertainty about incoming revenue and its contribution to his cash flow and by facilitating his ability to predict cost. If the manufacturer knows in advance what his long-term sales volume will be, he or she can plan fixed cost investments in production capacity (including the labor force) that approach optimal production rates to counter the rate effect.

With multi-year contracts, the manufacturer can enter into smarter business arrangements with sub-contractors for higher quantities of material or longer-term agreements for just-in-time delivery. The end result is a lower total cost for production that the manufacturer passes on, at least in part, to DoD. If the Pentagon cancels the contract before completion, there are usually penalties it pays to reimburse the manufacturer for his up-front investments to support the contract. The principal advantage of multi-year contracts is better program stability. Their disadvantage to DoD is the loss in programming and budgeting flexibility from year to year. The disadvantage to Congress is a perceived loss of control over funding, or at least a reduced opportunity to debate the need for the program in successive years.

For the F-22 program, the Air Force is trying two new incentive programs to create cost reductions to stay within Congress-mandated spending caps, without reducing from 295 the overall number of aircraft they purchase. By spending $475M over the next five years, the Air Force anticipates avoiding the currently projected $2B program overrun, thus, they expect a re-

7. The frigates' limitations were exacerbated when they were assigned to aircraft carrier battle group operations. The *Perry* class was conceived and designed to be convoy, replenishment ship, and amphibious group escorts which made their slower speed and less extensive combat suites understandable design decisions.

turn of ten dollars in savings for each one invested now. Half of this seed money is from slowing the F-22 production schedule in the early years and half is already in the production budget for contractor cost-saving incentives. The Air Force will implement incentives in two forms: up to 70% of savings from target costs will be returned to contractors and the Air Force will consider paying outright for a contractor's capital improvements, e.g., new machinery, training or software for advanced fabrication techniques if they create a substantial overall cost savings.

From this discussion, one can easily see why DoD values competitive bidding so highly when it issues contracts and why the consolidation of the defense industrial base has become a cause for concern. Fewer companies competing for contracts translates into higher costs for DoD and, in the worst case, a single-source supplier can name its own price. Currently, only one shipyard in the U.S. can build and overhaul nuclear-powered surface ships so there is only one place where DoD can turn to build nuclear-powered aircraft carriers. The Navy does not have very much leverage over cost—there is no other bidder—so all that remains is to provide performance incentives to reward efficiency on what are fundamentally cost-plus contracts based on costs calculated by the manufacturer. This is not illogical: nuclear-powered aircraft carriers are sufficiently unique and expensive to make competition impractical because there is simply not enough work to keep two shipyards open.

On the other hand, preserving both submarine construction shipyards was the most compelling reason for building the third *Seawolf*-class attack submarine; the network of sub-contractors with their specialized skills would have disappeared before there was enough work to again support a second shipyard. Re-establishing the skills base for a second shipyard would greatly increase the cost of the new attack submarine when it goes into full production. In fact, Congress, DoD, and the Navy carefully distributed the work on the new Virginia-class attack submarines to protect the existence of both shipyards and to maintain at least some form of competition. Similarly, there is great concern whether the loser of the Joint Strike Fighter competition will stay in the military aircraft business.

How much defense industrial consolidation the government should allow and whether DoD should in effect subsidize competitive sources by the way it performs acquisition are contentious issues. If we accept that DoD is going to need unique and expensive weapons systems manufactured by privately-owned companies, we must accept that we are going to pay for them. Whether we can obtain more effectiveness by allowing additional market-driven consolidation that reduces contactor overhead costs or through competition, albeit somewhat artificial at times, is unclear. Our instincts tell us DoD will save money in the short term through consolidation but not in the long term as the number of sole-source suppliers increases and, as a side effect, the competitive spur for innovation is diminished.

CASE STUDY: THE ANALYSIS PHASE–COST
USMC MEDIUM-LIFT REQUIREMENTS: THE V-22 OSPREY AND HELICOPTERS

To measure cost, the Institute for Defense Analyses developed fiscal data for two sets of Marine medium-lift aircraft fleets, each with a 20-year aircraft life cycle. As we discussed at the end of Chapter 3, the first set of alternative fleets (Cost Level I) was based on the Marines' previously stated lift requirements for the assault elements of three brigades. For these larger fleets, DoD would purchase a fleet of 502 V-22s or an equivalent capacity in helicopters for the Marine Corps. The second set of smaller fleets (Cost Level II) was based on the capital investment the Department of the Navy announced it was willing to make when it canceled the V-22 in favor of a replacement helicopter fleet.

The Level I 20-year life cycle cost was $33B (all dollar figures in this case are FY88 constant dollars) and the second-level fleets were based on $24B, the funding level the Department of the Navy was willing to budget to buy a helicopter fleet to support Marine Corps missions. This $24B would buy 356 V-22s. Thus, IDA fixed the 20-year life cycle cost at two levels—$33B and $24B—and then examined the effectiveness of aircraft fleets for each funding level.

Although the mission requirements IDA studied supported the Marines long-term goal of over-the-horizon amphibious assaults, Secretary of Defense Cheney was particularly concerned about near-term (FY91-97) costs in the upcoming Future-Years Defense Program, particularly for his next budget. The Level II costs for the V-22 in this period were $13.1B while the helicopter alternatives ranged from $5.2B to $11.7B. When IDA slowed the production rate of V-22s and delayed full operational capability by two years, the near-term V-22 cost decreased to $7.7B, bringing the V-22 in line with the helicopter alternatives. IDA used relevant costs, e.g., they included the remaining research and development cost for new aircraft and displayed sunk costs before FY91 without incorporating them in future cash flows.

IDA computed the projected aircraft cost for each model with the DoD standard Aircraft RePricing Model and included initial spare parts. They calculated cash flows by multiplying aircraft costs by the annual production rate, including 100 additional aircraft for the Navy and Air Force for the V-22s. IDA based helicopter-operating expenses on the Department of the Navy's Naval Rotary Wing Aircraft Operating and Support Cost-Estimating Model. Since the V-22 is not a helicopter, IDA blended maintenance and component re-work costs from the Navy's Fixed-Wing Model. (All these models are mathematical models that generate cost estimates based on systems characteristics like weight and speed.) IDA's results for Level II are reproduced below:[8]

AIRCRAFT ALTERNATIVE	COST INCURRED FY 1991–97	NET PRESENT VALUE	YEAR BRIGADE ASSAULT
V-22 Nominal Production	**13.1**	**16.3**	**1996**
V-22 Slowed Production	**7.1**	**13.0**	**1998**
New Helicopter	**6.6–8.7**	**11.8–13.0**	**1999**
CH-47M	**5.8–7.9**	**11.6–12.8**	**1997**
CH-60(S)/CH-53E	**8.4–10.5**	**13.6–14.8**	**1996**
CH46E+53E	**8.3–10.4**	**13.3–14.5**	**1998**
Puma/CH-53E	**9.0–11.1**	**13.6–14.7**	**1998**
EH-101/CH-53E	**9.6–11.7**	**14.0–15.2**	**1997**
President's FY90 Budget	**5.2**		

*** Includes new Marine medium and heavy-lift aircraft and 100 Air Force and Navy V-22 variants**

8. Table 12 from the IDA Study (SECRET) *Assessment of Alternatives for the V-22 Assault Aircraft Program (U)* by L. Dean Simmons (Alexandria, VA: IDA, June 1990). This table is unclassified.

The economic analysis of the V-22 aircraft and the helicopter alternatives was high in validity. IDA used FY 88 constant dollars, adjusted for net present value,[9] to compare the two fleets in a manner consistent with our course concepts—they measured the right kind of cost, in this case near-term cost and life cycle cost. The most important characteristics we look for while assessing reliability are accuracy and whether the results can be replicated. The economic models IDA used provide consistent answers over a wide spectrum of choices, i.e., they are useful for assessing more than the six helicopters and the V-22 alternative fleets. Naturally, contractors tend to be optimistic about their cost forecasts (another circumstance beyond IDA's control) that favor the undeveloped aircraft as well as the V-22. Overall, IDA achieved very good levels of reliability in their economic analysis. The discussions that followed the study's release did not challenge the cost estimates, a convincing indicator of solid economic analysis.

IDA scored well in practicality. They used existing data wherever possible and they were as careful as they could be using estimates. Better data on an unproven technology like the tilt-rotor was simply not available in 1990. They could have enhanced some of their cost estimates by using more than one option to explore a range of cost estimates for the V-22. Overall, we give IDA high marks on the economic analysis in their study.

Summary

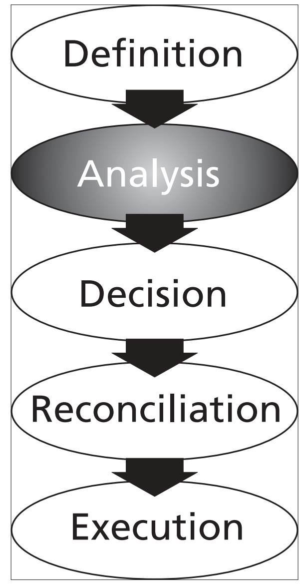
Cost is the measure of how many resources we will consume to implement an alternative. It is an essential part of almost every analysis of DoD program and policy options. Increasing effectiveness incurs greater costs: while proponents may talk about benefits and opponents may emphasize cost, they are both actually talking about cost and effectiveness. There are many types of cost and we are most interested in future costs that are relevant to our organization; we discourage executive decision makers from dwelling over sunk costs. In order to compare alternatives across different years, we convert current dollars to constant dollars to remove the effects of inflation.

Executive decision makers must seriously consider the ramifications of reducing short-term costs by reducing procurement rates: they increase average cost, increase total program cost, and delay replacements thereby increasing the age of equipment in the operating forces. The cumulative effect of slowed or reduced procurement cannot be eliminated without direct compensating action—increased production rates and more procurement funding unless we cut force structure.

9. Net Present Value is an accounting technique that attempts to capture the time-value of money beyond inflation which, as we discussed earlier, is the seemingly inevitable rise in prices of goods and services. Assuming the prices are equivalent, one would rather have a television set now rather than ten years from now, i.e., an object purchased in the near future is more valuable to the user than one purchased later. Net Present Value is calculated mathematically like interest or inflation using a discount rate, a rate set by economists to express this future value. We always apply the discount rate to constant (inflation-adjusted) dollars. The discount rate is by its very nature contentious, varies between experts and organizations, and is carefully guarded by business planners because it is key as they decide between investment strategies. Net Present Value calculations are required for all DoD acquisition programs using discount rates set by the President's Office of Management and Budget. For this study, the discount rate was fixed at 10 percent.

C H A P T E R 5

ANALYSIS CONCEPTS: UNCERTAINTY AND RISK



The habit of gambling contrary to reasonable calculations
is a military vice which, as the pages of history reveal,
has ruined more armies than any other cause.
—B.H. Liddell Hart: *Thoughts on War,* 1944

IDEALLY, WE WOULD LIKE to approach any decision fully able to predict its outcome with certainty. As we have said, during analysis we identify alternative courses of action to achieve our decision objective and compare them based on our criteria. Put another way, we ask ourselves what are the consequences of each alternative in terms of cost and effectiveness? And, what happens when we cannot predict these consequences with certainty? Depending upon the magnitude of our uncertainty and the importance of the uncertain issues, our ability to make a sound decision may be reduced substantially or eliminated entirely.

Taken to the extreme, when we have no sure knowledge, i.e., a situation of complete uncertainty, we may as well decide by chance. In most cases, this is undesirable—although there are certain decisions for which a coin flip is as good a method as any other. Uncertainty, in our framework, is the amount of doubt that we have about our projections of cost and effectiveness.

Risk rises from our uncertainty. Risk is the possibility of failure and of suffering loss or harm because of our unsure knowledge. The loss or harm is tangible and we can predict its effect although we may be less confident whether or not the effect itself will occur. In this chapter, we address the sources of uncertainty and risk and we present some methods for coping with them.

Objective Probabilities

Uncertainty and risk are very much about what is knowable and what is not. When we can identify a set of outcomes, when we are confident we know their behaviors in terms of cost and effectiveness, and we can predict the likelihood of any particular outcome statistically, we have objective probabilities. We can build objective probabilities whenever we can tabulate data from what has happened in a large number of prior cases that are similar to our present decision. Commonplace examples include weather forecasts and baseball batting averages. Each is calculated based on probabilistic information about what has happened over time with a set of initial conditions. Therefore, we can be confident of the likelihood we assign to different outcomes when the next occasion arises, so long as we believe that the future and the past will behave in a similar manner.

Weapons performance characteristics are a good military example of objective probabilities. When weaponeers calculate a missile's circular error probability vis-à-vis the aim point, they are establishing a typical objective probability-based measure. We cannot predict whether any particular missile will hit its target, but we can speak knowledgeably about the likelihood of a hit, as well as misses of various distances. In logistics, we measure the performance of an inventory system – and its possibility of failure—with objective probabilities. No inventory system is good enough to provide completely accurate information about every item in its database. But we do know with great confidence the likelihood of an inventory error. Precisely the same is true for quality control measures. While we cannot know whether an individual munition will explode, based on tests and experiments we can know how likely it is to be a dud and adjust our targeting plans and inventories accordingly.

Subjective Probabilities

In conditions of true uncertainty, one or both of the following conditions exist: we cannot predict the consequences of our alternatives with confidence or we cannot know their probabilities of occurrence. This happens when we are confronted with decisions that involve unique elements or at least elements too dissimilar to support a statistical probability based on the past. For example, suppose we must estimate the likelihood of a baseball batter getting a hit who has never batted before. We have no basis for assigning an objective probability to that estimate.

Our interactions with other people frequently fall into this category. As we grow to know someone, we can usually begin to see broad patterns in his or her behavior. However, people, situations, outside events, etc., change so continuously that we are seldom in a position to say that a certain individual has a 30 percent probability of doing any particular thing and mean it literally. At least, that expression of probability would have a very different level of meaning than a 30 percent estimate attached to a weather forecast or a gambling bet. Because these predictions are largely intuitive, they are subjective probabilities.

Consider the range of defense problems that fall into this second category in which uncertainty and risk are important elements. All our estimates of international behavior are based on our limited knowledge of the plans and intentions of others. We cannot say that Iraq has a .5 likelihood of launching an attack in the same sense that we can say that a typical 155mm artillery shell has a .9 probability of exploding. The U.S. plans military forces based on the two overlapping major theater war requirement. It would be enormously helpful if we could estimate the chance that a second major theater war might actually occur as we fight the first. We know that a second war is possible, but beyond that we cannot say. If we could say more, we could know much better whether or not planning for two overlapping major theater wars is a good use of our defense resources.

In the same sense, we cannot predict the likelihood that a Kosovo-type conflict or another crisis will occur in the upcoming year. If we could, we would have a much better understanding of what level of preparedness we require for such conflicts and force planning would be a science instead of an art. Because we do not know the probability of another or several similar events that will require deployment of U.S. forces, we must be ready for these scenarios whether their chances of happening are high or low.

Strategic planning and decision making are rife with uncertainty and the consequences of dealing with uncertainty are significant. Even at the operational level, can we know by objective probability how an adversary will respond to one battle plan versus another? For example, which course of action is more likely to produce an adversary's surrender: air attack or ground attack? Does the surrender of one adversary to one type of attack represent a universal truth or an exception? Does a new solution to an old problem represent new truth about the possibilities of the future or a fleeting aberration?

SILENT AIR ASSAULTS: GLIDERS, THE "OTHER" AIRBORNE

In 1940, during World War II, the Germans were the first ever to employ gliders in airborne assaults, allowing troops to land ready to fight and with unit integrity, at least at the squad level. The glider-delivered troops did not require parachute jump training and gliders could carry heavier equipment, like jeeps and anti-tank guns, than aircrews could shove out of aircraft in flight or land with a parachute. Also, because gliders could be released far from their destinations, the tow planes' motors did not alert defenders, whereas the parachutists' air transports were audible to them. Both parachutists and gliders were very vulnerable to ground fire and required open fields for landings. Because of their tactical advantages, many force planners thought glider troops would supplant paratroopers in airborne assaults. In their initial use in combat, German gliders landed engineers atop Fort Eban-Emael in Belgium and achieved complete tactical surprise; the strongest fort along Germany's western border fell quickly, unlocking supply lines into the Low Countries and France for the Panzers that had advanced through the Ardennes Forest.

After this success, how would you have reacted as a force planner in 1940, forced to choose between forming paratrooper or glider regiments? Which unit is more cost-effective? Allied and Axis planners hedged; their armies built both airborne and glider forces. It turned out the glider had a short life in combat, from 1940 to 1945. Casualties among gliders, glider pilots, troops, and cargos were high and for the most part they consumed material and manpower resources that could have been better allocated elsewhere, e.g., a pilot with the skill to land an unpowered aircraft on an unfamiliar, unimproved field at night might be better employed flying as his or her primary duty. A glider pilot without that skill was a hazard. Almost all major contemporary militaries still include paratroopers.

Should this result have been foreseeable? Could the force planners of 1940 have removed the uncertainty in their decision by doing more research or experimentation? Or was the glider an appropriate weapon for its albeit brief combat life? These same issues confront force planners today as the services strive toward Joint Vision 2020: should we press forward with new organizations and structures or use this "strategic pause" to do more experimentation and reduce uncertainty before committing to new, expensive paths?

If we could resolve issues about the efficacy of weapons or the effectiveness of air power with certainty, we could easily decide the current debate about the relative division of labor—and therefore resources—between air and ground forces. Logically, we suspect that striking certain kinds of targets with air power will incline an adversary to become conciliatory. But in any particular case, we cannot tell a Commander-In-Chief how likely it is that an adversary will indeed react this way.

Yet there is a helpful way we can characterize uncertainty. For instance, we may say about the baseball batter for whom we have no data, "He will not get a hit." By that, we mean that we rate his chances of hitting at less than 50 percent. This estimate of .5 is not an objective probability because no information exists about his previous batting performance. But there may be observations we can use to build a more refined estimate of his chances of hitting during his first at bat. We may evaluate the way he swings the bat or stands at the plate. We may see whether he appears confident or hesitant. While these clues are not the basis for an objective probability, they may support an expression of how likely *we believe* he is to get a hit. We generate a probability without data, without knowing the past (in a scientific sense), instead we evaluate our state of mind and determine how confident we are that the batter will or will not get a hit. We create a subjective probability.

We use subjective probabilities all the time in defense decision making to express our evaluations of uncertainty. For example, early in the Cuban Missile Crisis, President Kennedy estimated the chances of war between the U.S. and Soviet Russia as one in three. Subjective probabilities are less likely to be expressed numerically than objective probabilities; in fact, expressing them in numeric parlance can lead to misinterpretation. We believe the chances of two major theater wars overlapping one another are more than trivial; but when we say the chances of war are one in three, we do so knowing the chances of war are not measurable statistically like a batting average. Rather, this probability-phrased expression is shorthand for articulating a subjective estimate of likelihood based on experience.

In other words, we may not be able to calculate the statistical or objective probability of an event, but we still may have an idea for some reason of its likelihood. That idea may come from experiences which, although not identical, we believe are relevant to the probability we are trying to estimate. The estimate may also come from expert knowledge or intelligence we have about the specific circumstances of the event. Plainly put, the subjective probability does not measure the frequency with which something occurs, it captures how confident we are that something will or will not happen. This is important because as we analyze options and sort information provided by others, we need to know which data is subjective and which is objectively derived; the former is far more open to interpretation and dispute. In some cases, we may justify our assumptions based on subjective probabilities. When we do, we must be sure to inform our decision maker and be ready to be challenged by other stakeholders during reconciliation.

UNCERTAINTY AND FORCE PLANNING IN EUROPE

In 1994, the U.S. European Command knew it would continue to draw down from its 350,000 European Troop Strength (ETS). It also knew that the bulk of these troops would be Army and Air Force units stationed in Western Europe. ETS does not include rotational naval forces in the Sixth Fleet, typically comprising an aircraft carrier task group, submarines, patrol planes, and the Mediterranean Amphibious Readiness Group with its embarked Marine Expeditionary Unit-Special Operations Capable.

The force reduction process began opportunistically when the U.S. Army's VII Corps deployed for the Gulf War then re-deployed to the United States instead of returning to Europe. But, after their departure, how much deeper, if at all, should ETS have been cut? A General Accounting Office study looked at alternative force structures from 150,000 to 25,000 European Troop Strength in 25,000-person increments with some estimates on the influence and capability of each size force.

In 1994, the Commander-In-Chief of the U.S. European Command confronted a changing, and in many ways uncertain, mission. Nonetheless, he had to submit a force structure architecture for his command to the Secretary of Defense, the President, and Congress soon after the Gulf War, cognizant that many political leaders expected a peace dividend in part demonstrable by lowering ETS. The CinC defined the problem for his staff by using the National Military Strategy and focusing on the largest variable, the Army component, and by identifying several boundaries:

- Overall ETS would be more or less 100,000.
- The U.S. would retain leadership of NATO and that implied an assemblage of ground forces that equated to an army corps.
- U.S. forces would need to respond to several different kinds of crises, probably nearly simultaneously.
- Interest in peacetime engagement and the exercise tempo would increase, with outreach programs to Central Europe and more activity in Africa and the Mediterranean Sea than before.

By definition, an army corps has at least two divisions and a set of supporting forces. A standard army corps, however, would consume most of the 100,000 European Troop Strength goal by itself. The force planners therefore needed to see how much of the corps actually had to be stationed in Europe. Reviewing the regional situation and the Defense Planning Guidance, they identified four likely near-simultaneous scenarios requiring U.S. ground forces: (1) a peace operation in Former Yugoslavia (one division); (2) a peace operation in the Middle East (one brigade); (3) a humanitarian disaster (one battalion plus many support units), and; (4) a non-combatant evacuation operation in Africa (Southern European Task Force plus aviation elements). The forces not committed to crisis response would meet the exercise and engagement commitments.

The result was that the two divisions in V Corps in Germany have only two of their three brigades in Germany. Also, only about two-thirds of the corps-level units, those chosen for their utility in crisis response, are stationed in Europe. The resulting force, then known as the "Credible Corps," was sufficient to sustain U.S. leadership of NATO's European command. V Corps troops in Europe, plus independent and Echelon-Above-Corps Army support units, total 65,000. The Air Force sized their tactical force to support the Credible Corps and to support what is now Operation Northern Watch, which has been flown from Turkey into Iraq since 1992 (and flown primarily by rotational units from outside the theater). They packaged their logistic force for immediate reaction to the crisis scenarios, arriving at 35,000 Air Force personnel in theater. The Navy, already essentially headquarters and logistic organizations, maintained 14,000 people in Europe for a total ETS of 114,000 that was within the CinC's goal.

How well did the CINC and force planners assess uncertainty in 1994? They did pretty well. In 1996, the European Command deployed the two brigades of the 1st Armored Division to Bosnia as part of the Dayton Accords Implementation Force and the Air Force flew over Bosnia in Operation Deny Flight and later in support of the peace operations. The lack of a peace settlement in the Middle East forestalled a new peacekeeper deployment there, but the continuous presence of Task Force Able Sentry in Macedonia absorbed the forces planned for the Golan Heights. In 1996, the U.S. responded to the humanitarian crisis in Rwanda with a large airlift operation (requiring force protection) and evacuated several African embassies. Typically, the U.S. European Command oversaw six Joint Task Forces in 1996, a pattern that continues today.

That said, what the planners did not and could not foresee was the open-ended nature of many of these commitments. The U.S. European force structure, capped near 100,000, is not deep enough to support *rotational* deployments for these deployments. As a result, force planners are reviewing ETS to decide again how much force structure the U.S. needs to maintain in Europe to maintain its leadership of NATO and continue shaping the security environment, i.e., can European Troop Strength be safely reduced if we assign crisis response to CONUS-based general purpose forces? Do we still need a corps-equivalent in Europe? How much more benefit and influence with our Allies do we derive from permanently stationed versus deployed forces?

Risk and Uncertainty Profiles

Because of the inherent differences between objective and subjective probabilities, we approach each differently to minimize its detrimental effect on our decision making. First, however, we must carefully assess which parts of the decision contain risk and uncertainty, whether there is more knowledge we can gain, and what the consequences of those risks and uncertainties are. To do this, we develop a risk and uncertainty profile. It consists of the answers to these three questions:

- What precisely do we not know that we need to know to make a decision?
- How much more knowledge can we gain about them?
- What are the consequences of these risks and uncertainties and are they important?

By answering these questions—which parallel our considerations of validity, reliability, and practicality for evaluating criteria—we categorize the unknowns and decide which are worth our further attention and whether that attention will pay off. Usually, these answers center on another examination of our criteria, an extension of our earlier validity, reliability, and practicality evaluations.

The first question above is about validity: what do we need to know vice what information is at hand, regardless of how easily we can obtain information? Left to their own devices, many analysts will provide us with that which they can expeditiously collect; we seek instead to identify an ideal—what we need if perfect and limitless information were available. After we identify what we want to know, we examine how much is knowable about each criteria and how difficult it will be to collect more knowledge—another look at reliability and the quality of our measurements and data. For that which is knowable we seek to build objective probabilities. For the other unknowns, we will or must settle for subjective probabilities.

To decide how important an item of risk or uncertainty is to our decision, we can use a technique called sensitivity analysis (explained more fully in Chapter 7). Sensitivity analysis allows us to assess the potential impact of each risk and uncertainty on the outcome of a decision by examining the results of each alternative in isolation. In sensitivity analysis, we vary the effect of a single risk or uncertainty over what we believe is their plausible range of values while holding everything else constant, and then we examine the various results. If the overall outcomes do not vary greatly from one another, the decision is not sensitive to that risk or uncertainty; if the outcome does vary greatly, then the decision is sensitive to that unknown. In other words, we ask ourselves: How bad or well could a risky or uncertain event turn out to be, and does that matter to our decision? Knowing this determines whether we need or desire further investigation, a measure of practicality we apply to our analysis.

For example, suppose we are trying to decide how to arm strike aircraft for a mission with multiple targets of varying importance. Targeteering data tells us which weapons perform best against which targets in terms of a probability of kill. We also have statistics on weapons reliability. In short, we do not know how each weapon will actually perform, but we know the objective probabilities of each weapon against each target and therefore the risk of failure of any particular weapon. We have other uncertainties: how many aircraft will reach their weapons launch points? How many weapons on how many planes shall we dedicate to the highest priority target? Too few and the target may survive the strike; too many and we may have to fly another strike against the secondary targets that we could have destroyed during the first mission. By assessing the objective probabilities associated with munitions, we decide how many weapons we need to destroy the target; then, after assessing the subjective probabilities, we decide how many aircraft and how many more weapons beyond the earlier number we will assign to the strike.

Strike planners do this analysis by combining the databases and their experience. What they seek to uncover is whether, throughout its plausible range, any particular unknown matters greatly in this decision. They focus their energy upon those risks and uncertainties that are important with regard to the objective. In a decision where even relatively small amounts of error may matter a good deal, a risk or uncertainty deserves great attention. In a decision where being generally correct is good enough, only unknowns with large impacts are of further interest to us.

Additionally, creating the risk and uncertainty profile is necessary because acquiring more knowledge about important risks and uncertainties consumes resources. This raises cost-to-benefit and practicality issues. Do the resources we dedicate yield enough new knowledge about risk or uncertainty to improve our decision and therefore justify their expense? By culling the important unknowns from those less so, we avoid wasting resources on issues with little impact, and, when resources are limited, we can prioritize intelligently.

A RISK AND UNCERTAINTY PROFILE FOR NATIONAL MISSILE DEFENSE

The risks and uncertainties currently surrounding National Missile Defense (NMD) are being hotly debated. What would a risk and uncertainty profile include when applied to NMD and how is the Department of Defense likely to manage the attendant unknowns?

The risk of failure includes virtually all the design and engineering aspects of NMD – especially technological risk. The probabilities of detection of incoming missiles, especially discrimination from decoys by the ground-based radars and the kill vehicles' infrared sensors are physically knowable and we will discover them through simulation, testing, and experiments. We will also determine the probability of kill by the interceptor once an incoming warhead is detected and tracked. In the same way, the objective probabilities of various types of failures can be assessed and specified. The program managers will likely manage these risks by modifying designs for maximum cost-effectiveness and then buying out the risks that remain by procuring enough interceptors to ensure that, considering the objective probabilities of failure along each step from detect to engagement, the aggregate risk of overall mission failure is reduced to an acceptable level.

But there is another category of factors for which our knowledge is uncertain and not quantifiable. For example, how large an attack should we prepare for? How much warning are we likely to have? Will the attacker use sophisticated tactics? Will the attacker use penetration aids like decoys and chaff? These are unknowns for which we cannot calculate objective probabilities. Their likelihood depends upon future choices by our adversaries. For this reason, they are uncertainties with critical implications for the NMD force structure. We likely will attack these unknowns by first learning as much as we can to narrow the areas of our ignorance.

Good intelligence is key. How many weapons of what kind do our potential enemies possess? What kind of command and control doctrine does each adversary use and what is their likely salvo doctrine? What steps must they take prior to launching missiles and what are the signatures from those steps that will provide us with warnings? This intelligence should narrow the range of possible attack sizes and sequencing to subjective probabilities.

Once intelligence has told us all it can, we can design a National Missile Defense force structure on the basis of further scenario analysis. We will construct a range of hypothetical attacks to determine the NMD force structure that best defeats the aggregate. The scenarios should include the worst possible case, the worst plausible case, and middle-of-the-road cases, e.g., attacks based on a terrorist (vice nation-state) attack. We can then compare the resulting national missile defense force structures on the basis of cost and effectiveness, allowing senior leaders to make informed decisions about the level of capability that they wish to fund.

Dealing With Risk and Uncertainty

There are three straightforward and popular ways of dealing with risk and uncertainty. The first is simply to continue to solve more of the unknowns, thereby reducing uncertainty. Next, we can acknowledge that there will always be some risk of failure for any alternative, and we can attempt to buy out some, or all, of this risk. Finally, if the risk cannot be bought out, we can compensate for it by adjusting the attractiveness of an alternative by incorporating risk into our calculations.

REDUCING UNCERTAINTY

Generally, when we have uncertainty we desire to calculate objective probabilities. If we can create them, then we will likely understand enough about cause and effect to know whether or not we can change those probabilities if we so desire. Improving the reliability of our objective probabilities to predict outcomes may simply require more research if the necessary data already exists but is not at hand. If the information does not exist, then we may need to conduct experiments in a laboratory or at a test range.

Whether or not we choose to invest the resources to define objective probabilities for an unknown returns us to the practicality issue: is the knowledge gained worth the resources consumed? The Navy sometimes shock tests a new ship to gauge the quality of its construction and its resilience to battle damage by detonating a large explosive charge near it underwater. The shock test is expensive and the hull flexing decreases the strength of the ship tested by making the hull more brittle (for the same reason one should buy a new motorcycle helmet after an accident). Is the knowledge gained worth the cost? Sometimes. Therefore, the Navy shock tests the lead ship of every large class and additional ships if they make major design changes. We improve our understanding by uncovering objective probabilities when we decide it is practical to gather the additional information we need. If not, then we continue to deal with unknown probabilities as uncertainties. Of course, the same logic prevails in the use of improving subjective probabilities, which will be discussed later in the chapter.

BUYING OUT RISK

Suppose we find that not all of our alternatives provide enough certainty in their outcomes. We may be able to use our second technique of risk reduction: we buy out some of the risk. We ask ourselves (or, more likely, our analysts) whether additional resources could reduce or eliminate the risk of failure and, if so, what is required? The answer is usually more money, time, or equipment. For example, once we know the objective probabilities, we can reduce the risk of a failed air strike by increasing the number of aircraft assigned or the number of weapons they expend. Both involve increasing resources. In effect, we convert risk into something else, in this case, weapons systems. Note that we have not changed the objective probabilities that any particular aircraft or weapon will accomplish the mission. But we have reduced, or bought out, the risk that the mission will fail by increasing the resources devoted to it based on our knowledge of those objective probabilities. This approach to managing risk is quite common and intuitive. We encounter engineering redundancy, another good example, all the time in our professional and personal lives.

Again, note that we buy out the risk by transforming risks into resource consumption. We can sometimes buy out the risk to compare different alternatives on a common basis. For example, suppose we need to choose one of three designs for a new aircraft program. Each alternative has families of risk, benefit, and cost criteria. For this illustration, we selected one representative criterion from each of these categories to illustrate buying out risk: probability of a major mechanical malfunction during a mission (risk), maximum speed (benefit), and unit price in constant dollars (cost).

Alternatives	Risk (mechanical failure)	Benefit (maximum speed)	Cost (production price)
A	0.20	Mach 1.5	$20 million
B	0.15	Mach 1.0	$25 million
C	0.05	Mach 0.8	$30 million

Table 5-1. Three Aircraft Alternatives with Unequal Risk

Based upon table 5-1, which aircraft is the most prudent purchase? We can eliminate some of the complexity of this decision by buying out the risk of a major malfunction associated with Design A and B and adding the resources needed to do that to each of their costs.

Alternatives	Risk (mechanical failure)	Benefit (maximum speed)	Cost (production price)
A	0.05	Mach 1.5	$32 million
B	0.05	Mach 1.0	$30 million
C	0.05	Mach 0.8	$30 million

Table 5-2. Three Aircraft Alternatives with Equal Risk

Table 5-2 enables us to compare designs more simply on the basis of cost and effectiveness (only) by translating risk into cost. In so doing, the least expensive alternative can become the most costly while an expensive alternative can, in the light of risk, become a bargain. In this example, we have elevated the mechanical reliability of aircraft A and B to the same level as aircraft

C. Note that their benefits—top speed—remain unchanged. Now our decision is reduced to whether we value the higher speed of aircraft A enough to pay $2 million more per copy than we would for aircraft B.

EXPECTED VALUES

This leads to the third procedure we use to deal with risk. Suppose all the alternatives involve significant risk, and that the cost of buying out the risks to equal levels is too high. Or, suppose our circumstances call for simply accepting risk and choosing the best overall alternative, even though their benefits vary and each carries a differing level of risk, e.g., one has the highest level of risk and the greatest effectiveness. For situations like this we use the Expected Value approach. Like buying out risk, it is a way of adjusting the attractiveness of an alternative to reflect its probability of success.

Expected value computations can become complex and we need not become fully conversant with the mathematics involved. But, because we may have to compare alternatives based on expected values (calculated by someone else), we will familiarize you with the basic concepts here. Expected value computations use the concept of utility that we discuss more fully in Chapter 6, "Combining Criteria." Utility provides a way of translating the different attributes of alternatives into a common unit of measure that reflects their usefulness or value with respect to the decision objective. We can quantify these values for each alternative under each criterion and sum them to make direct comparisons. To obtain the expected value of an alternative, we multiply its utility (cost or benefit) by its probability of occurrence (risk).

For example, suppose a lottery prize is worth one million dollars, and one has a 100 percent chance of winning the lottery—only one ticket will be sold. The expected value of this ticket is one million dollars ($1M times 1.0) minus the cost of the ticket. Anyone who paid more than $1M for this winning ticket was unwise. Now suppose each ticket has one chance in two million to win the one-million dollar prize and tickets cost one dollar apiece. Is a ticket a cost-effective purchase? Because the chance of winning is one in two million, the expected value of a ticket is fifty cents: the benefit ($1M) times the probability of winning (0.0000005). The lottery makes $.50 on every ticket sold. To be cost-effective for the ticket purchaser, the ticket would have had to cost less than $.50, but, of course, no lottery could stay in business on that basis.[1]

We apply the same principle of expected value in defense decision making. For example, suppose we must select a weapon system alternative. One has a utility of 50 if all the subsystems perform as specified, but there is a 30 percent risk that they will not. The other weapon system has a utility of 40, but with a risk of only 10 percent of subsystem failure. The first system has an expected value of 35 (50 times 0.7) and the second system has an expected value of 36 (40 times 0.9). The higher risk reduces the expected value of the more effective alternative below that of the less effective one. This is as far as we need go in understanding expected value. Be mindful that there is nothing magical about expected values—they simply combine benefit (or cost) and risk into a single convenient number.

Improving Subjective Probabilities

Uncertainties that we can express only in terms of subjective probabilities are more problematic to decision makers than those with objective probabilities. The fact that we cannot predict the

1. That is why "Prairie Home Companion" radio host Garrison Keillor refers to state lotteries as a tax on people who did not do well in high school mathematics.

alternatives' outcomes with certainty or that we cannot assign probabilities to those outcomes based on statistical knowledge has important implications. It is harder to think about buying out this kind of uncertainty because we do not know when we have committed enough resources to eliminate it. Nor can we use the expected value approach since no objective probability exists on which to base an expected value calculation. A variety of other approaches do exist to accommodate uncertainty and risk, but there is no escaping that uncertainty and risk limited to subjective probabilities is among the most difficult aspects of defense decision making.

BETTER INFORMATION

The first approach to reduce uncertainty is to acquire more information. Perhaps the uncertainty we face is due, at least to some extent, to ignorance that we can dispel, if not to the point of objective probabilities. Perhaps we have not discovered everything we can. We should review our information about the problem and consider additional sources. This may be as simple as going to the library, searching the Internet, or making a telephone call; or it may require expensive research. As always, the issue is practicality: cost versus benefit. Do we have good reason to believe that more information has a reasonable chance of reducing the uncertainty? Is the decision deadline looming such that by the time we obtain the information, the decision will have become a moot point?

REFINE SUBJECTIVE PROBABILITIES

We often have genuine uncertainties that cannot be resolved by obtaining more information because what we need is not knowable. We cannot know the objective probability that there will be two overlapping major theater wars. We cannot know how close we actually came to nuclear war with the Soviets during the Cuban Missile Crisis. We cannot know statistically whether an assault on a hill will succeed. We cannot develop an objective probability that host-nation support for strategic mobility will be available as we plan for it. We cannot know how a particular unit will perform in combat as a function of its training. How then can we proceed to make rational defense decisions?

We can assign subjectively-derived probabilities to uncertainty as expressions of confidence based on our experience and the analysis we have done thus far. We think it likely that an Army unit will succeed better in combat if they complete a rotation at the National Training Center. That improvement we may describe subjectively in a change of readiness from C2 to C1 in unit readiness reporting or by increasing the unit firepower scores in a wargame. Subjective probability may be valuable because it provides a way to treat uncertainties somewhat like risk and it provides a frame of reference for discussion. This is advantageous because we have seen that risks can be relatively easy to incorporate into a decision. But we must keep in mind that subjective probabilities are prone to various kinds of errors that objective probabilities are not. We can compensate for these errors and guard against them, just as we do with the flaws in our memories, but we cannot be sure that our subjective probability estimates will be accurate. The most common errors are:

- ***Wishful thinking:*** We may estimate the subjective probabilities of various outcomes based on how desirable we regard those outcomes. But, of course, the likelihood of an event has no connection to how desirable we think it is. For example, defense decision makers who plan and execute an operation tend to be more optimistic about its chances of success than individuals uninvolved. The decision making prior to the Bay of Pigs invasion of Cuba in 1961 was distorted by this phenomenon. So was the decision making prior to the 1980 rescue attempt of the U.S. hostages held in Iran.

- ***Selective perception:*** We may not include all the factors that matter when we estimate subjective probability. In a similar vein to wishful thinking, we may include only those factors that we regard as special or are otherwise notable. For example, aviators may overestimate the impact of air strikes achieving a campaign's objectives, or dismiss the use of missiles or Special Forces as viable options for neutralizing a target.
- ***Experience:*** We may bias our subjective probability estimates of all outcomes based upon our memories of similar events. The more recent or powerful our memory is of similar events, the higher we will estimate the probability of the outcome that seems important or dramatic to us. Most people estimate that the chances of an airplane crash are higher than they would be otherwise if an airplane has crashed recently. For instance, in the wake of the September 11, 2001 terrorist attacks, many Americans are traveling by car instead of plane, even though historic experience shows air travel to be far safer. Perhaps this is also one reason why it is so often said that militaries prepare for the last war. Military veterans of the Vietnam War tend to be more opposed to peace operations than other constituencies.
- ***Framing effects:*** The way we define the question may significantly influence our subjective probability estimates. For example, we may change our estimate of the outcome of an operation depending upon whether we are asked to predict the probability of success or the probability of failure, even though the two estimates should be complementary.
- ***Overconfidence:*** Perhaps the most dangerous and prevalent influence is our sense of infallibility. We are generally far too confident of our ability to personally estimate the probability of an outcome with great accuracy. In a wide variety of scenarios, individuals are repeatedly much worse at making estimates of probability than they think they are. A great deal of sound research has repeatedly confirmed this disturbing problem.

There are two fundamental methods for improving our subjective estimates. The first is to be aware that virtually all of us are prone to making one or more of the above errors when we try to estimate subjective probabilities. By being aware of them, we can compensate for them.

Second, and usually more successfully, we can involve other people in our problem solving. While virtually all of us are prone to these perceptive errors, the forms they take in each of us are likely to be different. By involving several people to estimate the subjective probabilities of the same events, the weaknesses of one participant may be offset by the strengths of another. At some point, too many participants become unproductive. Our recommendation to involve others in your decision making may be difficult for those who prefer solitary reflection and have a low regard for group problem-solving activities. Despite these common and understandable sentiments, effectively including the military judgment of others is an important part of executive decision making. That said, some ways of obtaining group views are better than others, depending upon how much time is available and how much trouble we wish to take.

DELPHI METHODS

At the very least, our choice of decision-making participants should be based on their background, availability, reputation, and often the organization they represent. There are a variety of ways to bring their ideas together usefully. The most common, popular, and quickest approach is a BOGSAT: a Bunch of Guys Sitting Around Talking. There is not much more to add. We bring the right people together, usually after providing some read-ahead material, and moderate a discussion. It is helpful to have an agenda to guide the discussion and ensure that the essential issues receive attention. A recorder is a good idea, as well.

More elaborate is the Delphi[2] method. It begins by having the participants vote anonymously on a set of proposed subjective probabilities. While sometimes there will be tight convergence of opinions, usually there is a wide difference. Next, we discuss why each participant agrees with some estimates and disagrees with others. This exposes our assumptions and arguments to the group's assessment. Informed discussion will highlight when and how a participant may be making one of the misjudgments discussed above. Others can detect and correct those errors, and estimates will change in the process.

We follow the expository discussion with another vote. Usually, we see some convergence of opinions. Depending upon how much, we may have another conversation and another vote. At some point the group's estimates will stabilize to the point where any remaining possibility of convergence is not worth the effort to obtain it. Ultimately, we may obtain a consensus estimate or we may get two or three clusters of estimates. Occasionally, no convergence happens at all. In any case, we have important information about what people whose expertise and judgment we trust think about a critical risk or uncertainty. We learn what they believe are the cause and effect relationships that shape uncertainty, what assumptions they think carry important weight, and the direction the uncertain outcomes may take. In most cases, this information is more valuable than what you could have developed ruminating alone. Naturally, we can attain much of this knowledge, somewhat degraded, on a less formal basis.

Now that we have obtained subjective probabilities that are as informed as possible, we can begin to treat the unknown as if it had objective probabilities. For example, we can assess whether one alternative involves much less uncertainty (as expressed by subjective probability) than the others. If so, we may select that alternative to avoid uncertainty, if, at the same time, we can satisfy our minimum requirements for effectiveness and cost. Similarly, we can consider buying out the uncertainty. In this case, we can convene our group and ask them to make subjective probability estimates of how additional resources will affect the risks and uncertainties. Again, keep in mind that all these judgments are completely subjective. We cannot have the same confidence in our outcomes as we do when working with objective probabilities.

WORST-CASE SCENARIOS

Armed with subjective probabilities, we can buy a hedge against the most plausible and important range of outcomes, an investment against undue loss due to the uncertainties. One way to do this is by choosing alternatives with the flexibility to cover the range of outcomes that matter to us. But, like all capabilities, flexibility is not free. It may make wonderful sense to select an alternative that allows us to achieve our goals in a variety of circumstances, and the financial cost for this flexibility may be straightforward, but other costs may be subtler. For example, by preparing to respond to many situations, e.g., to achieve *Joint Vision 2020*'s Full Spectrum Dominance, we may not be particularly well trained for any. This is our concern for general-purpose forces like infantry battalions, multi-role fighters, and ships. Because we train them for many eventualities, our training costs go up while our readiness for warfighting may simultaneously

2. Delphi was the location of the Greek god Apollo's shrine where oracles tried to predict the future.

degrade. Although these kinds of cost may not be readily apparent, we must include them. We pay all costs, whether we know it or not.

Another classic way of using subjective probabilities is to select the alternative that is most effective against the worst plausible outcome, the worst-case approach. The crucial assumption we have to make, if we choose a worst-case alternative, is that all other plausible and important outcomes can be subsumed in the worst one. That is, if we can handle the worst case, then we must be able to handle less dire circumstances by definition. We know, however, by logic and experience that this assumption may not be true or may be only half-true.

During the Cold War, U.S. conventional forces focused on stopping a Warsaw Pact thrust across the inter-German border. The scenario that force planners envisioned required large-scale, high-intensity operations against the Warsaw Pact. Forces and concepts designed to stop an armored juggernaut in Europe were not well suited for the other applications that were originally thought to be less stressing. Vietnam is a case, or several cases, in point. Before the U.S. military's frustrating involvement, the French discovered this painful truth. The Soviets, of course, were effectively defeated in Afghanistan and later in Chechnya using forces meant for a war against NATO.

The lesson here is that the worst-case approach to handling uncertainty can be sensible and in some ways efficient, but we must take great care to be sure the worst-case assumptions are realistic and acknowledge when they do not transfer to other circumstances. When we have limited resources, the worst-case approach may form the basis for our force structure, but we need to include other capabilities when we know we will confront other circumstances. Our limited resources in the 1990s and our emphasis on the worst case (two overlapping major theater wars) has forced DoD planners to make just that kind of difficult decision and has resulted in today's High Demand/Low Density units, e.g., civil affairs, military police, tactical control elements, reconnaissance and air surveillance aircraft. Because of uncertainty about the future, they had to decide whether to accept more risk of failure in major theater wars by building more active duty combat support and service support units to support peace operations or whether to maintain the more traditional focus on warfighting. They chose the latter because the consequences of failure were so much higher even though the likelihood of peace operations was much higher.

EXPECTED VALUE WITH SUBJECTIVE PROBABILITIES

Finally, we can apply the expected value approach we used earlier to refine objective probabilities, now using subjective probabilities—cautiously. For example, most planners would say that the chance of nuclear war with the Russians is very low. That is a subjective probability. There is no objective or statistical way to know this since we have not had a series of nuclear war preconditions to tabulate. If we deem the probability so low, why do we spend all the resources that we do on nuclear forces? The answer is that, although the probability seems low, the consequences of being wrong are astronomical. This is an expression of expected value using a subjective probability. When we multiply the huge negative value (utility) of nuclear war by the small probability of its occurrence, the negative expected value is still far too large to ignore, so we continue spending resources on nuclear deterrence. Depending upon how confident we feel about that negative expected value, we may even gain some sense of how many resources are worth devoting to this mission.

We can follow a similar process with any other decision involving subjective probability. Our confidence in the resulting expected value depends upon how confident we feel in the judgments behind it and if the analyst does not tell us, we need to ask. Using these procedures, we can begin to unpack the problem of how many resources we should expend to be ready for a second overlapping major theater war. In a similar fashion, we can address how much effort we should expend to offset the uncertainty that host-nation mobility support will not be available.

ASSESSING RISK IN PREPARATION FOR THE QUADRENNIAL DEFENSE REVIEW

Since 1996, each new Presidential Administration during its first year in office has been required to report to Congress its defense strategy and the force structure and programs it requires to execute that strategy. The 1997 Quadrennial Defense Review (QDR) was roundly criticized for having a serious disconnect between its strategy and the force structure it identified to execute it. The latter was largely constrained by budget considerations and has been noticeably frayed trying to execute the former with a $50B per year funding shortfall.[3]

To facilitate rapid execution of the 2001QDR, the Chairman of the Joint Chiefs of Staff commissioned a study group of four field grade officers, one from each service, led by Michele A. Flourney at the National Defense University (NDU) to identify issues and options for the new administration as it conducts the QDR.[4] The group observed that within DoD, we do not have a commonly accepted risk management framework.

They proposed a structured methodology for examining risk. The NDU team proposed that decision makers evaluate what they called the strategic military risk of each force structure alternative; in other words, they proposed examining each force structure alternative's efficacy executing a national military strategy (vice a national security strategy which employs diplomatic and economic tools as well). Risk, in their terms, was the risk of failure.

Ms. Flourney's group broke strategic military risk into three categories[5], each having two subcategories as shown below:

Operational risk is how well (or poorly) a force structure alternative achieved the current military strategy. Force performance is the U.S. military's ability to achieve military objectives in support of war plans and peace operations; force sustainability is how well the military maintains its readiness over time across the spectrum of conflict, from engagement and presence to humanitarian operations, for peace operations and crisis response, and through major theater wars.

Force preparation risk is how successfully (or poorly) the military prepares for future operations, based primarily on future force structure and doctrinal choices and procurement strategies. Transformation risk refers to force structure designs for the most likely scenarios while hedging risk concerns less likely but still possible scenarios like a resurgent, expansionist Russia that threatens NATO and forces a return to Cold War practices.

Affordability risk evaluates whether the force planning choices DoD makes are affordable, first concerning the allocation of resources within DoD and then considering DoD's portion of the over-

3. Congressional Budget Office, *Budgeting for Defense: Maintaining Today's Forces* (Washington, D.C.: Government Printing Office, Sep. 2000).
4. Michele A. Flourney, Report of the National Defense University Quadrennial Defense Review 2001 Working Group (Washington, D.C.: National Defense University, Nov. 2000), 8. The remainder of the descriptions of the groups work, including illustrations, are derived from pp. 31-32, 49-52.
5. The actual Quadrennial Defense Review Report (30 September 2001, Washington D.C.) ultimately settled on four dimensions of risk: Force Management Risk, Operational Risk, Future Challenges Risk, and Institutional Risk. However, it (in essense) dealt with risk in the same fashion as the Flournoy Group.

all federal budget. By evaluating and then aggregating all six elements, one can derive an overall sense of how much strategic military risk a strategy funded at a given level of resources entails.

The study group makes four points about this process: (1) it is compatible with many different models; (2) sets aside contentious considerations about national will (casualties) and leaves that to decision makers outside the force planning realm; (3) the lack of reliable peace operations models hampers the analysis of force structure alternatives and their efficacy in various low-end scenarios, and; (4) the lack of reliable models for peace operations will hamper the QDR, but nonetheless the QDR must set a general strategic direction for peace operations to close the strategy-to-resources gap.

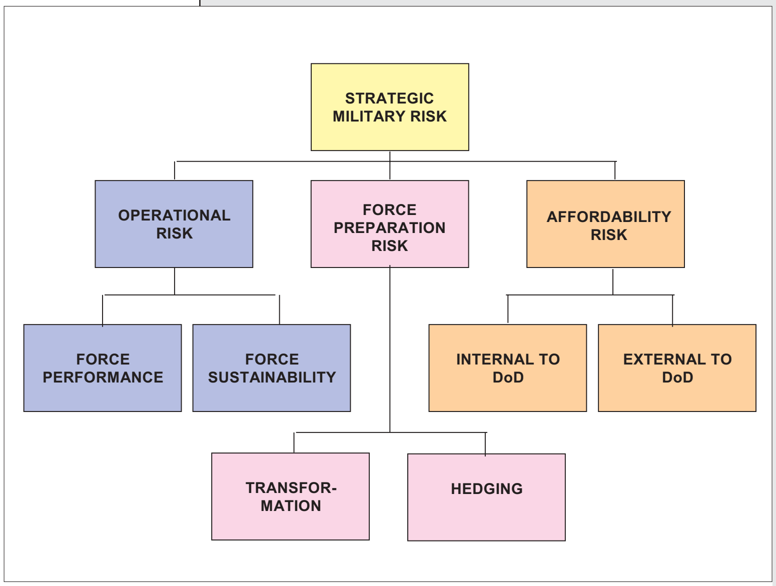


Figure 5-1. Measures of Risk Wiring Diagram Example

Each of the first four types of risk must be explored, they write, using the force-on-force analysis methods we describe in Chapter 8. After the force structure required to meet each strategy-driven situation is identified, planners can derive a force structure and its cost. Presumably, a zero-risk force would be fully funded and be able to accomplish all its goals simultaneously and almost instantaneously. Since that level of resources is very unlikely, they conclude, some forces will multi-tasked (as are general purpose forces today) and we may accept decreased response times by moving some capability to the reserve components. Funding may not materialize at the level DoD identifies (our current situation).

Resolving those tensions introduces risk, some of which we are living with already: the risk that there may not be enough forces simultaneously available for the maximum number of operations; the risk forces may not be available quickly enough; the risk that a high operations tempo will degrade sustainability; or the risk that force structure will not be funded adequately to transform and hedge. By studying and gaming the impacts of accepting risk in different areas, we can more intelligently decide where to accept it and plan for it, rather than watch it happen.

When applied to the QDR, the NDU group emphasizes that this risk analysis must be highly iterative before it will yield a worthwhile set of force structure alternatives. They provide a step-by-step process for each assessment in an appendix to their report; but before the process can begin, it needs a strategy with prioritized objectives. The prioritization implies where to take the risk. The study group also defines four levels of assessing risk:

- Low: failure is unlikely and the resources and time to achieve objectives is acceptable.
- Moderate: failure is unlikely, but achieving objectives will take longer and consume more resources.
- High: failure is possible but unlikely and more resources and time will be required.
- Unacceptable: failure is likely despite using high levels of resources and a lengthy timeframe.

The authors would have planners apply these ratings across each of the six types of risks above and they specify up to four criteria for each, e.g., analysts would assess force performance regarding protection of key terrain, time to achieve objectives, effects upon the enemy, and number of friendly casualties in scenarios across the spectrum of conflict. The worst level of risk assigned to a subordinate category or objective becomes the assessment for that risk, i.e., if a force structure's performance in a major theater war is low risk in the terrain, effects on the enemy, and casualties categories but high in terms of the length of the campaign, the risk under force performance for that force structure is high. Therefore, it is also high for the overall strategic military risk.

How useful is this methodology? It certainly captures the essential elements of force planning. Used as the NDU group did, risk is the counterpart of capability (the likelihood of success) and this process is very much like a bottom-up or capabilities-based force planning method. Defining risk as low, moderate, high, and unacceptable is helpful if they become standard in DoD. As the authors point out, the terms too often mean different things to different audiences. What is less clear to us is whether it is necessary or desirable to aggregate the different kinds of risk beyond the three—or even six—categories. Since they are calculated in different ways, using different criteria, evaluating a force structure as "moderate in strategic military risk" is not particularly meaningful compared to knowing it has moderate risk in affordability and low operational and force preparation risk. Likewise, two force structure alternatives may have equal strategic military risk for starkly different and meaningful reasons.

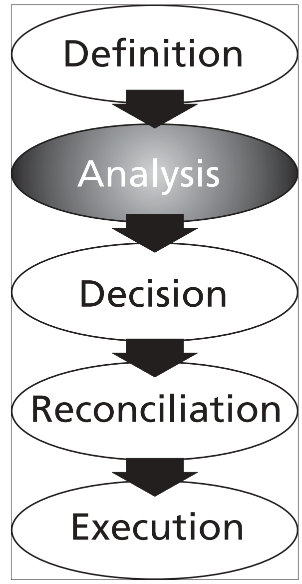
Notice, too, the subjective nature of many of the assessments: how much longer need a force conducting an operation take to slide its force performance risk from low to moderate? We can create some objective probabilities for some of these risk assessments, particularly affordability, but certainly not all. Here we are (in our lexicon) dealing with uncertainty and it will by necessity figure largely in this kind of "risk" assessment.

Summary

Uncertainty and risk are important aspects of defense decision making. Uncertainty results from our doubts about how much we know and risk stems from the possibility of failure that results from uncertainty. We evaluate them during decision making based upon how much we know about our alternatives. We create objective probabilities when we know, or can know, the statistical likelihood of an outcome. When we cannot arrive at definite probabilities for outcomes, we create subjective probabilities based on our best judgment.

The first method for coping with uncertainty and risk is to acquire more information about the alternatives by doing more measurement. We may improve our comparisons among alternatives by buying out risk—expending more resources to translate the risk into something else, like cost—if we have confidence in our objective or subjective probabilities. We can also address uncertainty and risk by using an expected values approach wherein we adjust the attractiveness of an alternative's cost and benefits by tying them to its probability of success. Finally, we may reduce uncertainty further by involving other people's expertise in our decision making to get their views on the information we have available.

Risk and uncertainty are ever-present in defense decision making. Indeed, they often dominate it. Our senior leaders make many major, high-level decisions despite distressing levels of uncertainty. As we get closer to procurement and operational matters, we tend to deal increasingly more with objective probabilities; strategic choices are invariably clouded by differing evaluations of subjective probabilities. Executive decision makers must understand both as they proceed through the Analysis Phase.



ANALYSIS CONCEPTS: COMBINING CRITERIA

Science is built up with facts, as a house is with stones.
But a collection of facts is no more a science
than a heap of stones is a house.
-Jules Henri Poincaré: *La Science et l'Hypothèse,* 1908

NOW THAT WE HAVE DESCRIBED the important considerations about alternatives we wish to measure, we will take the next step and combine them to facilitate our comparisons. Methods for doing this range from simple to complex, depending on the nature of the problem. In this chapter, we will begin by making evaluations based on a single criterion and finish the chapter with more complex weighted models that incorporate as many criteria as our resources allow.

Combining Cost and Effectiveness

The heart of the Analysis Phase is the evaluation and comparison of alternatives. We confront several challenges combining cost and effectiveness that vary with the individual decision. For example, DoD recognizes the tension between cost and effectiveness and deliberately reflects that tension in its organization. The Combatant Commanders (CinCs) focus on warfighting capabilities (effectiveness) while the services must balance those requirements against programs and overall force structure (cost).

We know that the least costly and most effective alternatives are rarely the same, and each is attractive to different organizations. The least costly options are naturally appealing to our political leaders and headquarters staffs who are charged to conserve the public treasure. They rightly seek to fund as many programs as possible to support all the services' needs. On the other hand, military operators are biased toward the most effective system, regardless of cost, as they seek to win battles quickly, with the fewest casualties. However, an overly operational preference in programming can lead to excess capability (gold-plating) that drains resources from other urgent programs. Executive decision makers, then, are forced to make tradeoffs between cost and effectiveness. We strive to present those tradeoffs in structured terms to make these choices clear.

We cannot always compare the alternatives fairly using the same measures of cost, especially when procurement options are in different stages of completion. For instance, we may be comparing three missile alternatives to counter a new threat: (1) upgrading a missile already in the inventory; (2) procuring a prototype missile undergoing flight tests, for which research and development is complete; and (3) funding a missile proposal on the drawing board. Comparing the costs of these missiles equitably is difficult. The first missile will have by far the lowest unit cost in terms of procurement; however, an upgraded missile will not last as long as a new missile and it has a very limited potential for growth. The missile on the drawing board will have the longest life-cycle cost measured from the present. How shall we identify the time horizon for calculating life cycle costs for this procurement decision? Use the shortest? Or the longest? Instead, should we use the procurement price of each missile in constant dollars? Do we really care if costs are compared equitably? How we frame the measures of cost may determine the outcome of the analysis before we make any further comparisons.

We encounter a similar problem with comparisons based on effectiveness, especially when multi-mission capabilities are at issue. (Rarely do our alternatives achieve the same levels of effectiveness for each criteria.) Take the comparison of the effectiveness of the new C-17 transport aircraft with the C-5, C-141, and C-130. The C-5 and C-141 are used for inter-theater lift; the C-130's mission supports intra-theater lift. The C-17 was designed to do both. The measures of effectiveness the Air Force uses to compare the aircraft must include both airlift missions, and the alternative fleets must each accomplish both missions. However, effectiveness alone is not the issue; we must take cost of alternative fleets in consideration as well.

Fixing Cost or Effectiveness

One technique for combining cost and effectiveness is to fix cost or to fix effectiveness.[1] When we fix one of them, we compare our alternatives on the basis of the other. Thus, when we select a specific type of cost as the single type (or measure) of cost that we will compare our alternatives against, we are fixing cost. Of course, our challenge when fixing cost or effectiveness is to ensure the alternatives are truly equal in performance or value in the area we fix. We have a myriad of ways to measure cost. If we select just one, then we ignore the others which may vary significantly, e.g., if we fix life cycle cost, we only measure that type of cost; if we fix near-term cost, we eliminate downstream operations and maintenance costs from consideration.

The same problem occurs when we fix effectiveness; we have to select one (and only one) measure of effectiveness and then select the lowest cost option. For example, what single measure is the best way to measure the effectiveness of a tactical combat aircraft? In 1996, in their evaluation of the affordability of the Defense Department's tactical aviation procurement plan, the Congressional Budget Office selected aircraft age, represented by the proxy measure Technology Generation, and thereby fixed the effectiveness of the F/A-18E/F Super Hornet, the F-22 Raptor, and the Joint Strike Fighter at the same benchmark. When DoD identifies a set of specifications for a system and sends out a Request For Proposal to contractors, it is in effect fixing effectiveness and preparing to select the alternative with the lowest cost.

Cost and effectiveness are inextricably related to one another; however, the relationship is rarely the same for each alternative. For example, contractors may include different support packages in their proposals. Their unit price may vary with the quantity procured, creating step functions in the unit cost profile as they open another production facility. In such cases, we may

1. If cost and effectiveness are both equal, then the decision maker has no economic or performance basis for making a choice and can select any of the alternatives that meet minimum requirements.

not be able to fix cost or effectiveness easily. When we cannot fix cost or effectiveness, we might combine them to help us choose between the alternatives.

Cost and Effectiveness Ratios

One straightforward method for combining cost and effectiveness involves constructing a ratio. To do so, we must isolate one measure of cost and one measure of effectiveness that each represents our range of alternatives. By selecting one and only one MOC, we are effectively saying that the other costs are irrelevant to our decision; likewise, as we select an MOE we establish that there is an MOE that dominates all the others in importance.

The principal advantage of the cost-effectiveness ratio is its simplicity; this is also its principal disadvantage. Again, to use ratios, we must be confident the single MOC and single MOE we select truly dominate cost and effectiveness. We can compare alternatives easily and quickly using ratios displayed as a simple number (like cost per square foot), on graphs, or as tables. Again, this technique breaks down quickly as the complexity of the problem increases and a single MOC or MOE no longer dominates.

Weighted Models

One of the most powerful ways we can combine multiple criteria is by using weighted models. Basically, we arrange our criteria into a hierarchical model and evaluate the importance of each measure to the decision maker. We then assign weights to each criterion that reflect that relevance. After we build the overall hierarchical model, we evaluate a range of possible scores for each criterion using a utility scale. The utility scale is based on how much value we place on an alternative's improved performance (beyond the minimum requirement) for that criterion. Finally, we evaluate each alternative using the complete model, multiplying each utility score times the weight for that criterion then summing to create a total score for each alternative. We will describe each step in further detail below.

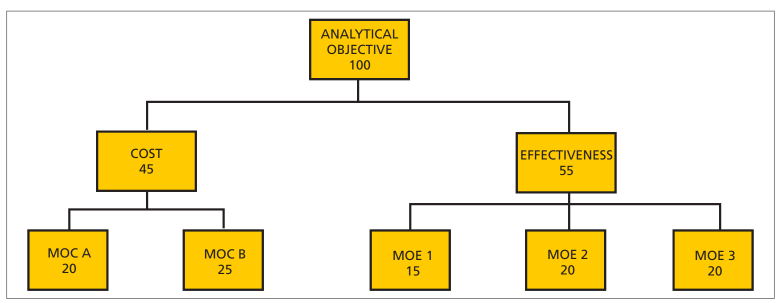


Figure 6-1. Weighted Model for Cost and Effectiveness.

Figure 6-1 is a typical weighted model with two primary categories of criteria, cost and effectiveness. The weights of a model may add up to 100, they may be normalized to total 1.0, or, if the model is built from the bottom up, they may combine to whatever number is the sum of the weights. Notice in figure 6-1 how the full weight descends from the analytic objective into successive tiers: first, Cost and Effectiveness, and then down to each of their subordinate measures. The weights of a subordinate tier must equal the weight of the parent tier above it.

Assigning the weights is a subjective process. It may be a group process, but it is the culmination of the experience of the people making the model. The manner in which we award weights may determine the outcome of the analysis, so we must have them approved by the decision maker and we may expect to have them challenged during the Reconciliation Phase. Thus, at this juncture, we are blending the mechanics of analysis with some very important professional judgments to make rational decisions.

UTILITY

After we design the structure of the weighted model and assign the weights to each criterion, we must decide how to assess our alternatives based on the lower-most criteria. This requires us to apply utility to the range of measures which defines our criteria. For example, if speed is one criterion, we must establish the range of speed for which we are interested and decide how important each incremental change is. We do this by constructing utility curves.

Utility is simply a way to show the decision maker's value-gained or lost by changes in the measurement of our criterion. Then, depending on where our alternative falls on the utility scale, we compare it with others. Conversion of all measures of cost and effectiveness to utility allows us to multiply values by the weights of criteria and convert numerous measures to one aggregate number for each alternative.

Utility must be dimensionless because we are going to add utility scores together. We derive utility scores from measurements and assessments that often do have dimensions, e.g., dollars, speed, range, man-days. To make the conversion from the direct assessment to utility scores to use in the model, we create utility curves that show value as a function of performance, cost, and sometimes risk.

As you might imagine, the key to this process is the care with which each alternative's performance on each measure is converted into the common currency of utility. Using a fighter aircraft example, suppose we wish to compare two alternatives. Aircraft **A** has a top speed of Mach 1; aircraft **B** has a top speed of Mach 2. If we determine that Mach 2 has twice the utility of Mach 1 for a fighter, the utility score of aircraft **B**'s speed should be twice that of aircraft **A**. If Mach 2 is only 10 percent more useful, then aircraft **B**'s speed-utility score should be 10 percent greater than that of aircraft **A**. You can see that by mapping speed in regard to its utility, you can generate a utility curve—a picture of the utility of each speed.

We can do the same for every criterion; every measurement or assessment falls along a utility curve that expresses its usefulness or value to our organization and the decision maker. The very act of identifying utility benefits our organization because the participants must have (or gain) a thorough and universal understanding of our core values and missions as we establish the utility curve of value versus performance. We can evaluate utility in a variety of valuable and creative manners, numerical and otherwise.

UTILITY DISPLAYS

Graphs are the most common method for displaying utility. A graph is a simple model that shows pictorially the relationship between two sets of numbers. Since we are translating the measures of our criteria into utility, graphs are especially powerful for converting performance and cost data into utility by showing how changes in performance or cost relates to value or usefulness.

To build a utility curve, first we establish a range of values for the criterion and identify the range of likely values for the set of alternatives. If we are selecting among armored fighting vehicles, one of our criteria may be Maximum Off-Road Speed. We set a minimum requirement (threshold) of 30 MPH to ensure the new vehicle can keep up with our current armored vehicles and those that we project will be in the inventory during the new vehicle's service life. We would like a top speed (objective) of 60 MPH to enable it to dash from point to point. These are the endpoints for our range of values for speed: 30-60 MPH that we will put on the horizontal axis as shown in figure 6-2.

Next, we decide how much resolution we want to distinguish performance levels among the alternatives: our options include every MPH difference (or even less), 5 MPH blocks, 10 MPH blocks, etc. After discussion within our organization and with the analyst, and with the approval of the decision maker, we opt for 5 MPH increments because they are significant enough to differentiate between alternatives without being too detailed. These blocks are often called bins. Each bin, in this example, has five values that we score equally, i.e., to us, there is no more value in a vehicle that travels at 45 MPH than there is for one that travels at 41 MPH.[2] This is another crucial intersection of experience and analysis. Our experience, and that of the operators, subjectively determines the utility of different levels of performance, just as it did when we established weights.

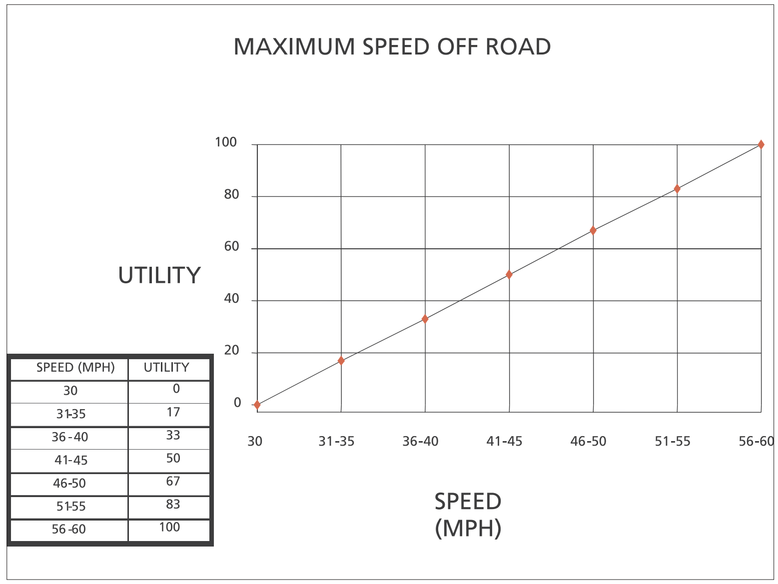


Figure 6-2. Straight-Line Utility Curve.

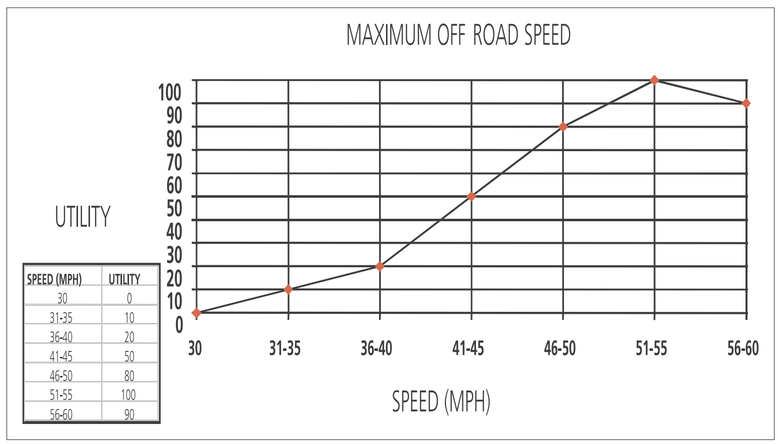


Figure 6-3. Curvilinear Utility Function.

For the vertical axis, we select an arbitrary scale of Utiles (unit-less measures of utility) from 0 to 100. We now build a data table by asking ourselves—or others, such as operators in the field—how much they value speed above 30 MPH, in 5 MPH increments as shown in the lower left corner of figure 6-3. If every increase in speed has the same value as the previous increment and the next, we will get a straight-line utility function as we show in figure 6-2.

Linear utility curves are good representations for many of the criteria we examine. They accurately describe cost profiles in which one additional dollar is worth the same to the decision maker as was the last dollar. There are, however, many circumstances that lead us to complicate our curves to better reflect reality.

Continuing the example above, we will probably find while discussing off-road speed with the operators that one 5 MPH increment is *not* as good as the next. For tactical vehicles, a maximum speed of 30 MPH is acceptable but not very desirable. Most troops desire speeds of at least 40 MPH, and many find speeds above 50 MPH significantly preferable. Towards the upper range of speeds, however, there is a leveling off and then a decrease in utility because operators never really need speeds over 55 MPH—and speeds beyond 55 MPH increase accident rates (according to our sampling). Revising our data table, we generate the curve in figure 6-3. Notice

2. If we are using a computer-based system, we could use the formula for the line instead of making bins. In the case in figure 6-3, Utility = 3.33 * speed (MPH) - 100; any speeds below 30 MPH result in negative numbers, another way of saying that alternative does not meet our minimum requirements.

that the bins remained in five MPH increments, but their utility does not change as evenly as before in the data table. The shape of the curve reflects our values; some incremental improvements are worth more than others.

Both the curves in figures 6-2 and 6-3 are smooth and continuous (without breaks). Sometimes we have utility curves that have abrupt changes in value (discontinuities) that create breaks in functions. This happens when a change in the criterion's value influences another criterion or triggers an event, such as when we receive a price break for buying in larger quantities.

Sometimes cost exhibits this behavior, e.g., if we budgeted a certain amount of money for a project, the utility of cost will decline smoothly as cost rises—lower cost has higher utility so our curve has a negative slope—until we reach our budget limit, marked by the star in figure 6-4. If we choose an alternative that exceeds the budget, the decision maker will have to obtain more funds, a distasteful but not impossible proposition. The utility of cost becomes linear again, once we cross that increased budget point, because each one-dollar is once more worth the same as the next. It is the act of having to get more money that creates the huge drop in utility; once we pass this hurdle, we return to the original slope where each one-dollar increase is worth the same utility as the next.

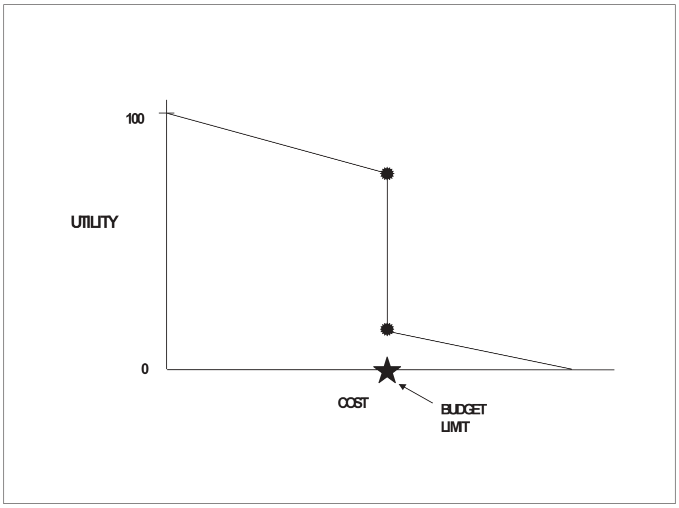


Figure 6-4. Discontinuous Utility Curve.

In the preceding examples, we have used quantifiable data for making our value assessments. While the utility of a particular value may be debated between decision makers, we can usually agree on a general function that describes rational decision makers' preferences. With qualitative criteria, however, we find ourselves marrying subjective traits (the value to the decision maker) and value judgments (the subjective data) on both axes of the utility curve.

We have several other options for compiling and displaying qualitative data. We can use a carefully defined descriptive scale, like the common traffic light scheme of green, yellow, and red for good, fair, and bad (see the next box). We may use qualitative terms from Very Low to Very High with any number of gradations in between. We can display our assessments directly in these same terms or, if we are going to incorporate them into a weighted model, we can convert them into utiles or numerical values, e.g., we convert Very High to 10 points; Medium to 5 points, and Very Low to 0 points.

ASSESSING OBJECTIVES: U.S. EUROPEAN COMMAND'S THEATER OBJECTIVES AND MEASURES

Figure 6-5 is the mechanism used by the U.S. European Command to display its quarterly progress towards achieving the Commander-In-Chief's Theater Objectives.[3] In 1995, then-CINCEUR General George A. Joulwan, U.S. Army, established nine theater objectives. He pronounced they were meaningless unless his staff could measure and report his command's progress toward achieving them.

3. Typically, USEUCOM has nine theater objectives such as Quality of Life for the U.S. military community, Strengthening NATO, Supporting Middle East peace initiatives, etc. The fishbone mechanism is based on the cause and effect diagrams from Michael Brassard's *The Memory Jogger Plus* published by Goal/QPC, Methuen, Mass., rev. 1996.

Basically, in conjunction with the rest of its Theater Security Planning System, CINCEUR's headquarters staff consolidates inputs, called Indicators, from its components and subordinate commands to assess each criterion, in this case measures of effectiveness. The indicators are often numerical, e.g., the number of exercises scheduled compared to the number executed, and the number of exercises this year compared to last year. A decision maker (a Division Chief at the O-6 level) evaluates each MOE in terms of Satisfactory (green), Concern (yellow), or Unsatisfactory (red).[4] Note in this case the colors (rather than numbers) express utility. Translating the individual indicators into the common currency of utility enables the decision maker to easily understand the meaning of the indicators.

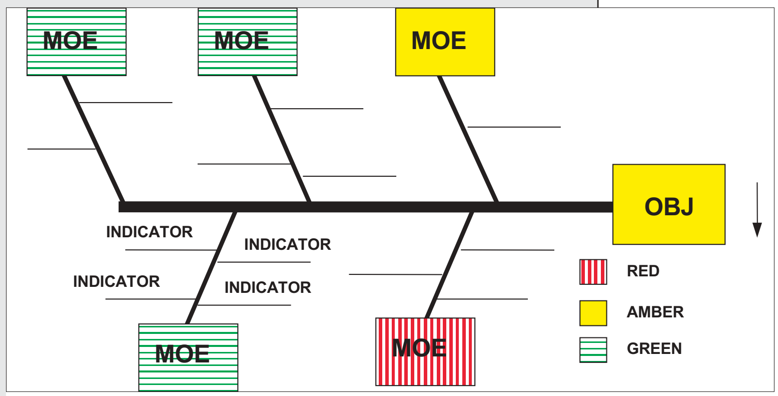


Figure 6-5: USEUCOM Red/Yellow/Green Measurement Chart

The Director (an O-8 decision maker) responsible for each objective makes an over-all assessment of the objective, assigns a color, and indicates an upward, downward, or neutral trend. The staff briefs the status charts to the CINC and his Component Commanders quarterly. Naturally, most of the discussion revolves around the areas that are not green; the briefing books include, behind each fishbone display, the MOE assessments and indicator data that support each rating.

After we have built utility curves for each criterion, we are ready to display them with our model. Recall that each criterion that ends a branch of the weighted model must be associated with a utility curve.

Figure 6-6 combines a weighted model with its utility curves. Note some of the characteristics of the curves. Where cost is concerned, one dollar is worth the same as the next to us for both MOCs so we have the highest score (maximum utility) for the lowest possible cost and follow a straight line down to zero for the highest affordable cost. The utility curves under the MOEs illustrate a variety of shapes that reflect the varied nature of the particular measures. In each case, we decide how much value we associate with exceeding the minimum requirement for that measure and whether one incremental change has the same value as the next. Thus the first MOE uses a curve, the second is a step function (yes or no), and the third uses bins to convert performance into utility.

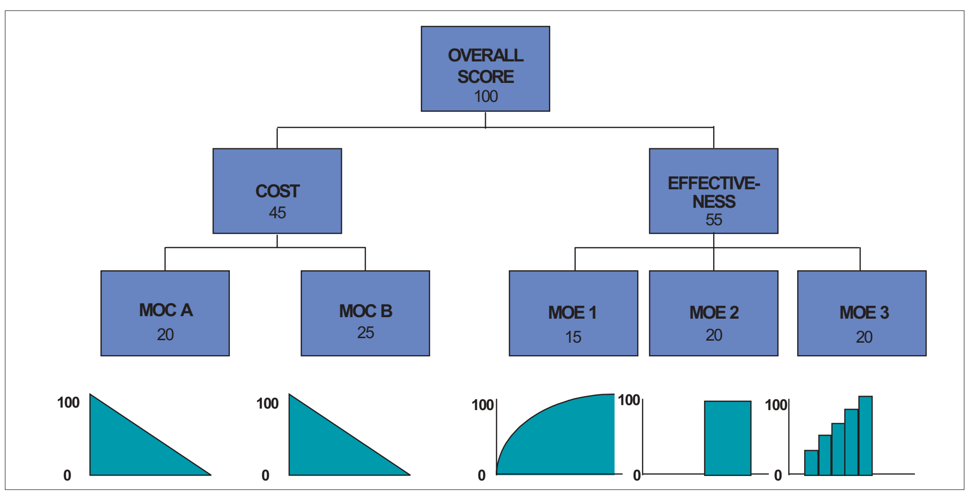


Figure 6-6: Model with Weights and Utility Curves.

4. It is a real challenge to convince organizations that reporting bad news about actions beyond their control (a red rating) will not be held against them, e.g., if the Arabs and Israelis refuse to negotiate with each other, we do not blame the action officers in J-5 Plans and Policy!

EVALUATING ALTERNATIVES USING A WEIGHTED MODEL

After we assign weights and decide upon the utility curve for each criterion, we evaluate the alternatives using the model, one criterion at a time. We arrive at a score for each alternative for each criterion; the score is its utility, which we multiply by the weight of the criterion as we show below.

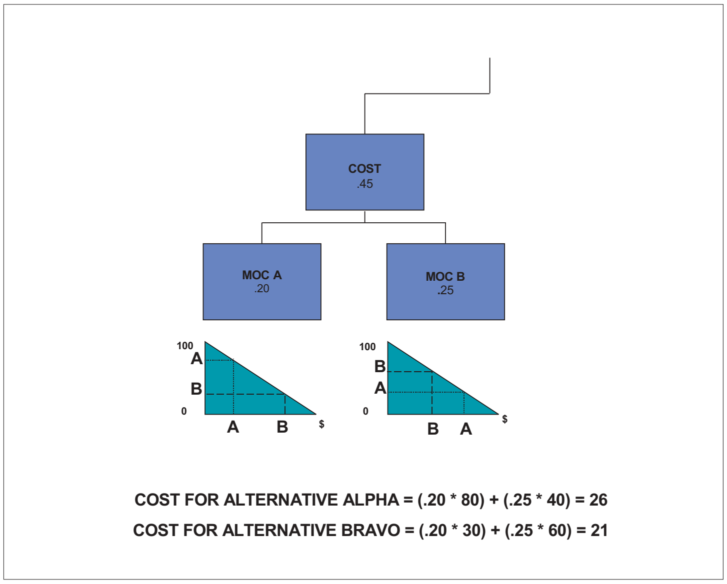


Figure 6-7: Calculating an Alternative's Score for Two Criteria.

To calculate the score, first we find its performance measurement or our assessment along the horizontal axis, and then we read upward until we intercept the utility curve. At the intersection, we read across to the left to find the utility value, as we show in figure 6-7, e.g., alternative Alpha has a utility value of 80 for MOC A. Then we multiply the utility value (80) times the weight (.20) to arrive at the score for this criterion: 16 for alternative Alpha for MOC A. In the figure, alternative Alpha scores 26 of 45 possible points for Cost while Alternative Bravo scores 21. We repeat this process for every measure at the end of every branch and then sum the scores to arrive at a total score for each alternative. In reality, the analyst will do most of this stubby pencil drill, but whether he or she uses a pencil or a Cray computer, the underlying principles are the same.

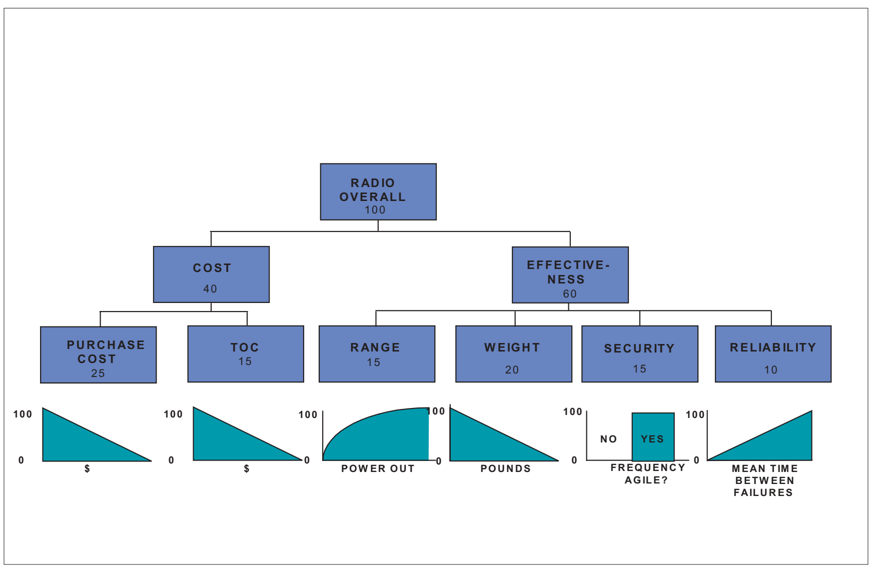


Figure 6-8: Weighted Model for a Portable Radio.

In figure 6-8, we have a weighted model and utility curves for a hypothetical portable radio requirement. The ideal alternative in a model like this would score 100; the worst (but still acceptable) would score 0; the vast majority of alternatives fall in between. The output from the model should allow us to identify preference and provide us with insights on how the alternatives compare and complement each other. These insights should include the strengths and weaknesses of each alternative and where the tradeoffs lie in selecting between options. After running the model, we should be able to see how we can modify a lesser-ranked alternative to improve its value to the decision maker. The results should also indicate where and how we could form hybrid alternatives to optimize the strengths of different options.

RADIO	PRICE	TOC	RANGE	WEIGHT	SECURITY	RELIABILITY	TOTAL
POPIEL 1995	23	15	0	20	0	0	58
WHAMMO 3000	11	7	12	16	15	6	67
ZONKER 101	4	14	15	6	15	10	64

Table 6-1. Radio Alternatives.

In table 6-1, we display the results of applying the model in figure 6-8 to three alternatives. Based on this model, the Whammo 3000 is the best choice—assuming that our weights correctly reflect the decision maker's perspective about what is important and that the utility curves reflect how we value differences in performance. Note, too, that the Whammo 3000 did not score highest for any single criteria.

Combining Risk and Uncertainty with Cost and Effectiveness

While combining cost and effectiveness is usually a straightforward process that becomes obvious from the Definition Phase and its analytic objectives, incorporating risk offers several options. As we discussed in Chapter 5, we can combine risk and cost by buying out risk and we can combine risk with performance by using expected values, multiplying probability times outcome. We can build our risk assessments into our models, or we may choose to evaluate risk separately after we evaluate our alternatives in terms of cost and effectiveness. The importance of including a consideration of risk and uncertainty is that it informs the Decision Maker as to what we don't know, as well as what we do know. Further, it helps him/her understand the negative consequences of each alternative. This more complete information package is what he/she needs to make the best informed decision possible.

CASE STUDY: THE ANALYSIS PHASE - COMBINING CRITERIA
USMC MEDIUM-LIFT REQUIREMENTS: THE V-22 OSPREY AND HELICOPTERS

The Institute for Defense Analyses used a fixed cost approach to conduct their V-22 and helicopter comparison. As we mentioned at the end of Chapters 3 and 4, they fixed cost at two levels in FY88 Dollars. First, at $33B, the cost they estimated for the Marine Corps desired fleet that would lift half of their assault force in the first wave, and, second at $24B that was the funding DoD was prepared to allocate for replacement helicopters. We reproduce the study's Table 4 again for easy reference:

Marine Corps Medium-Lift Assault Aircraft	Number at Cost Level I ($33B FY88)	Number at Cost Level II ($24B FY88)
V-22	502	356
New Helicopter	634	450
CH-47M	673	527
CH-60 (S)/CH-53E+	287/347	240/283
CH-46E+/CH-53E+	317/336	251/258
Puma/CH-53E+	330/322	260/246
EH-101/CH-53E+	252/335	200/256

Was this an appropriate methodology for this study? Cost was certainly a dominant aspect of this problem: indeed Secretary of Defense Cheney's overriding concern was the V-22's near-term cost in the face of many competing DoD programs. This study does a good job of demonstrating the value DoD received in return for its dollars, i.e., IDA can show, based on its assumptions (including those surrounding scenarios) and measures of effectiveness, that the V-22 generates more medium-lift per dollar. That is important and useful information. The only significant weakness of fixing cost in this manner is it removes the possibility that a less expensive helicopter fleet could accomplish the mission at a lower cost than $24B, however, that is unlikely because we can assume DoD explored that possibility when they identified their medium-lift proposal.

But what if IDA had fixed effectiveness instead? Naturally, they would have to focus on the assault scenario and the Marines would have to define a required build-up rate of combat power for the air assault some distance from the amphibious ships. The first challenge would be to identify a scenario; at this time, there was none on the shelf. The Marines were embracing Over-The-Horizon assault and just beginning to explore the ramifications of avoiding the build-up on the beach prior to moving to the objective, the sequence central to their earlier doctrine. Developing a new scenario, in the middle of this contentious medium-lift issue, was fraught with problems. The Marines (and IDA) would almost certainly have been accused of bias by V-22 skeptics whether justified or not—fixing effectiveness would expand the controversy rather than reduce it, as this study was intended to do, because stakeholders would argue about the scenario before they even got to the analysis.

There was no worthwhile cost-effectiveness ratio to create in this situation because there were two cost issues—near-term and life cycle cost—and Congress specified multiple effectiveness issues. IDA could have created a weighted model, but that may have added more complexity than was desirable given their target audience of DoD leadership and Congress. A weighted model would have been an amalgamation of what we already see: the two types of cost on one side and the eight scenario evaluations on the other. Assigning the weights would have been very contentious: imagine trying to reach agreement of the relative importance of cost and effectiveness. Given the circumstances, we think IDA chose the most appropriate methodology available.

Risk was incorporated in a very limited manner by IDA in its effectiveness evaluations and largely ignored as they considered cost. As events have shown, this was an important, albeit deliberate, omission. In the logistics sustainment scenario in particular, the effectiveness of the aircraft was based on mechanical reliability. To reduce weight and conserve space, the V-22 uses a flight control hydraulic system that produces 5,000 psi system pressure, more than double the pressure of most helicopters. Although IDA halved the time between failures the V-22 manufacturer projected, the doubled failure rate was still well below that of helicopters. With a new technology aircraft like the V-22, there was no way to create objective probabilities, however IDA's subjective probability was over-optimistic.

There was also a great deal of risk in the schedule and cost projections for the V-22 that fell outside IDA's study. One can argue whether it was their place to challenge Bell-Boeing's cost and schedule figures, but there was little doubt even in 1990 that they were optimistic. Essentially, they left those risks for the decision makers to tackle independently, probably for reasons of practicality. IDA was under severe time pressure to complete this study and by making assumptions and limiting it to what was measurable, they simplified its completion.

Summary

The methods we choose for combining effectiveness, cost, and risk criteria depend upon on the nature of the problem. We desire to make our comparisons as simple as possible by reducing the number of variables, which is most practical when we can fix either cost or effectiveness. If neither can be fixed but both have a dominating measure we can establish a cost-effectiveness ratio, either as a single number or displayed on a graph.

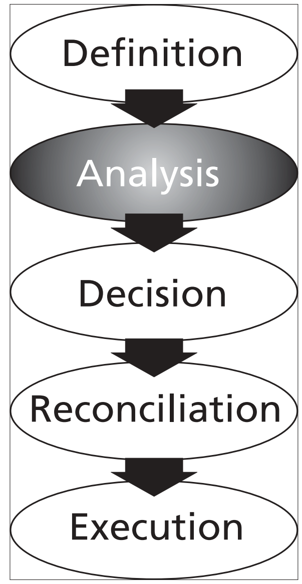
For more complex situations, we use weighted models that allow us to use as many criteria as time and money permit. As we construct these models, we in DoD establish the weights often without detailed foreknowledge of the alternatives. This leads to a more general model, but we may find, when we apply it to the alternatives, that some of our criteria are not very helpful discriminating between options because they all score equally in that area. In other circumstances, we may know the alternatives before we build the weighted model. This permits us to choose criteria we know will highlight differences, but the model may require revision if new alternatives are added that vary from the existing options in a new way.

For each criterion, we build a utility curve that we will use to translate measures and assessments into common, dimensionless units that reflect value or usefulness. We evaluate each alternative for each criterion using the utility curves. After multiplying each score by its weight, we sum them to reach a total score for each alternative. The data that we use to evaluate the alternatives, and to establish the utility scores, is largely beyond our control in the sense that it should represent the truth about each alternative, either by objective measure or careful subjective assessment. The results of the model can be affected by changing either the weights of the criteria, the shapes of the utility curves, or the values of the alternatives themselves. Because there are many subjective evaluations built into any weighted model, we will insist on knowing them when the analysts provide us the model.

We may incorporate risk with cost and effectiveness or we can study them in isolation after we have evaluated cost and effectiveness. However we decide to address them, we should make our intentions clear to the decision maker.

CHAPTER 7

ANALYSIS CONCEPTS: MODELING



Farming looks mighty easy when your plow is a pencil,
and you're a thousand miles from the cornfield.
-Dwight D. Eisenhower, speech in Peoria, Illinois, 25 September 1956

MODELS, SIMPLE OR COMPLEX, are the abstract constructs we use to compare alternatives. In defense resource allocation, models have four functions: organization of a problem, comparison of alternatives, measurement, and prediction. The first function we discussed in Chapter 2 as we defined and organized the problem. This chapter expands on the methods for combining criteria into models specifically designed to support analysis. The most frequent use of analytic models in DoD is to compare procurement and policy options on the basis of cost and effectiveness in the Analysis Phase.[1]

Our intention in this chapter is to familiarize you with the analytic modeling tools and terminology of the analyst, not to have you memorize classifications and characteristics of models. As you read this chapter, remember our goal is to make you a critical director and consumer of analysis who can confidently evaluate modeling proposals. By understanding the difference between good models and bad models and by subjecting analytic models to professional scrutiny in terms of validity, reliability, and practicality, you will be able to evaluate the quality of analysis without becoming a subject matter expert and thereby make good executive decisions.

Characteristics of Analytic Models

Analytic models are a specific class of models. They are so named because they are models composed of the separate parts of a problem—a problem identified by the analytic objective and the parts that were important enough to be facts, assumptions, or criteria. Analytic models require that their builders and users have an understanding of how those parts fit together. Analytic models are, at heart, based on the scientific method and they have a clear logical or mathematical structure.[2]

1. Within DoD, some cost and effectiveness analyses are given names that specify their structure and content. DoD uses the Analysis of Alternatives format, which superseded the Cost and Operational Effectiveness Analysis format in 1996, to support acquisition milestone reviews. DoD initiated the V-22 case study used throughout this text as a Cost and Operational Effectiveness Analysis.
2. Mathematics is often the language of modelers because of its wide applicability to seemingly unrelated problems. For example, the same form of equation describes the decay of a radioactive isotope, the swing of a pendulum, the decline of a population, etc.

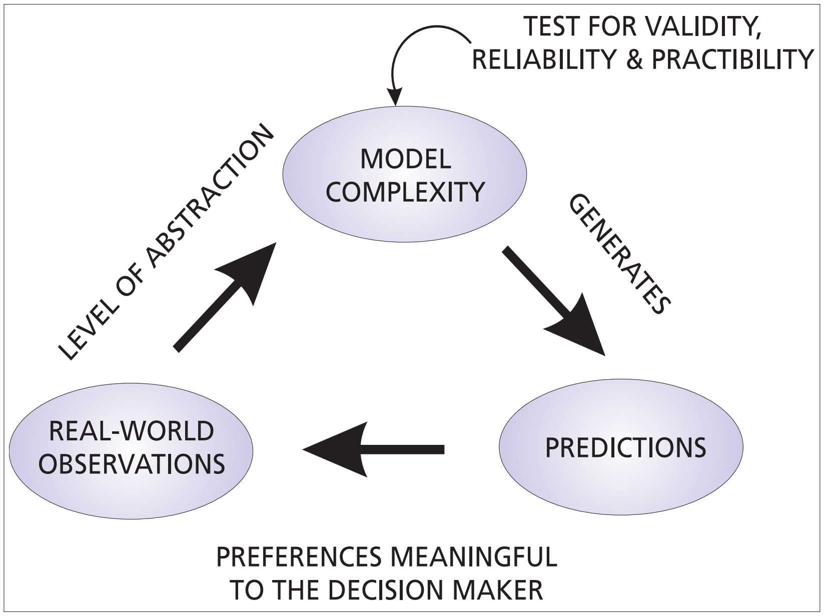


Figure 7-1. Models and the Scientific Method.[3]

With unlimited resources, including time and money, we would not need models, and we could satisfy our requirements with full-sized experiments and real-world observations. But we are forced to model by the prohibitive cost of experimenting in the real world, although the real world is our starting point for the analytic model as we show in figure 7-1. We will examine the major characteristics of analytic models—abstraction, complexity, and prediction—in more detail below. When we defined the problem, we set the stage for selecting many of the characteristics of our analytic models.

LEVEL OF ABSTRACTION

Our translation of the real world into an analytic model is an abstraction because it reflects a simplified reality containing only those factors and relationships we deemed important for solving the problem. The level of abstraction is inversely proportional to the degree to which the model literally replicates reality. Some models have a great physical resemblance to reality, e.g., mock-ups, prototypes, and miniatures. Others models, like the differential equations that represent the airflow around a ballistic missile warhead, bear little resemblance to the physical world they are used to investigate.

Most of the analytic models we use in DoD vary greatly from reality because they are based on mathematics or use scale representations such as time compression. Full-scale models, such as prototype aircraft, have few (if any) departures from reality. Policy analysis also takes advantage of full-scale models; before launching a new quality of life program, we usually test the policy in a pilot program as we see now being done with several military health program initiatives.

Iconic models are scaled down replicas of the real world, such as model airplanes, maps, globes, and photographs. We use iconic models to provide information without going to the expense or difficulty of building full-scale models, assuming the model's performance mimics real world performance. Note that the level of abstraction need not be connected to the complexity of the model. Some highly abstract models are very complex, such as those for space flight planning, and other highly abstract models are quite simple, like a flow chart.

COMPLEXITY

The complexity of analytic models is a function of the number of variables we need to measure, the resolution to which we measure them, and how many resources we have available to model—validity, reliability, and practicality concerns. We will also have some uncertainty about how well the interactions among elements of the model reflect reality. In the simplest case, we have rigid, full-scale analytic models; in the most complex models, we include interactions among the variables and insert events while the model is running.

3. Adapted from Samuel B. Richmond, *Operations Research for Management Decisions,* New York: The Ronald Press Co., 1968:.30.

Static models are representations of reality at a fixed point in time, freezing both time and position, such as a map or an organization chart. Because they do not incorporate change, they are generally simple and inexpensive. Their simplicity is helpful when we have a wide divergence of opinions and perceptions about the problem amongst the decision participants.

Dynamic models incorporate change in terms of time, events, and motion, e.g., a fuel usage curve that displays gallons consumed as a function of hours in operation, a graph of accidents as a function of crew rest, and computer simulations of weather conditions. Adding change to create dynamic models adds complexity and uncertainty to them. Dynamic models are often more difficult to describe and display, especially the interactions among variables. Even so, we use them to take into account interactions that we know to be important in the real world. To the extent that they succeed, dynamic models reflect the real world better than static models.

Yet, even simple static models may include important uncertainties; cartographers do not survey every square inch of the terrain represented by a map, yet every square inch is represented. Just as in the Definition Phase, the analysts and we are forced to make modeling assumptions to cope with uncertainty. We must take into account the uncertainty of important but uncontrollable variables in dynamic and interactive models, such as weather or the price of fuel.

NETWORKS, COMPLEXITY, AND UNCERTAINTY

Predictive modeling is based on the assumption that the future can, at least in part, be forecast by knowing the past and understanding how variables, including our criteria, act and interact within their environment. Some prominent theorists believe this assumption is fundamentally wrong. Chaos Theory is a well-known approach that describes a world where chance rules supreme and confounds our ability to predict outcomes that may vary wildly despite nearly identical initial conditions.[4] When we look at initial conditions and then outcomes long afterward, it is very difficult to identify exactly why the outcomes were so different. If, however, we start at the beginning and catalog the intervening events with ever more resolution, we can identify a linear series of decision points and chance occurrences (nodes) that keep branching out until we have a huge but exhaustive set of possible outcomes.

As we progress from node to node, some branches may merge into nodes with other branches, creating multiple paths to the same outcome—a network. The path we uncover by reverse engineering the outcome is one possible path among many in a network of unknown dimensions. Chaos theorists see any progression of events to an outcome as non-unique; one path along a network may be repeated later, but neither the path nor the outcome is predestined by the initial conditions. The longer the time interval and the more numerous the events, the larger and more complex the network and collection of paths and outcomes become and the more difficult it is to model. We can complicate the network further by adding more starting points.

Chaos Theory operates from the assumptions that: (1) the future is not linked to the past in a linear fashion, therefore we need higher order mathematics to approximate or model future behavior; (2) events in nature are very sensitive to initial conditions, therefore small, hardly measurable changes in one variable at the beginning of a chain of events can dramatically change the

4. A typical example posits a child dropping two ping pong balls into the Niagara River above the falls. One winds up washing ashore near the base of the falls and the other comes to rest on the coast of Africa.

overall outcome; and (3) stability in a chaotic system is unnatural and quite temporary, but where stability does exist it is determined by the relationship of very few variables. Therefore, controlling these key variables can control behavior in a chaotic system for a brief period. But, because this controllable time period is brief, we cannot predict the distant future with any degree of certainty.

The global weather system is an excellent example of a chaotic network. We cannot predict the weather accurately more than 72 hours in advance; our attempts to predict the weather further beyond the simplest generalities are futile, according to chaos theorists. But we know the probable range of outcomes from the global weather network, therefore civil engineers can plan their designs around 50-year storms, i.e., severe storms that statistically happen every 50 years, while no one tries to say exactly when the next one will occur. Also, we may be able to control the weather over a short period if we could identify and learn to manipulate the key parameters, such as by seeding clouds to precipitate rain.

These ideas can have important implications for the study of war. The network model is much more compatible with our experience of war than the chessboard. Analysts are not able to predict other than the grossest outcomes of war. If we can identify and learn to man-age the key determinants of the outcome of the process (which may be very few), then we can control the process of war over short, critical periods. This requires that we use higher mathematics and probability, accepting ranges of outcomes like worst case, best case, and most likely case to compensate for the much higher levels of uncertainty we will have to accept with network modeling.

PREDICTION

Analytic models make predictions about the outcomes we should expect; given our decision to use a particular model, our choice of input values, and our choices between alternative courses of action. If a decision-maker has confidence in a model and in the chosen set of input values, these predictions will help him choose a course of action.

Whenever we can, we evaluate a model's quality by comparing its predictions with real-world outcomes, then we calibrate it to better predict and improve our confidence in it. Of course, the extent to which we can do this depends on the kind of problem we are investigating. Certain problems make it relatively easy to test model results against real-world outcomes (e.g., how fuel consumption varies as a function of the kind of flight training we are doing).

The more the problem we are investigating involves predicting results in combat, the harder it will become to test model results against real-world outcomes. For one thing, we have a small number of real-world wars against which to compare our model results. In addition, careful historical analysis of actual battles shows that outcomes depend on a series of hard-to-replicate and unlikely-to-recur particular events.

Even if it's hard to know if a particular model is doing a good job of predicting combat outcomes, we can learn a great deal from modeling. For one thing, building a model forces us to say what premises we have to believe, in order to believe a particular prediction. Sometimes we can subject those premises to empirical tests. Depending on what those tests show, we can revise our prediction and, ideally, get closer to understanding "ground truth."

PREDICTION AND THE MOBILITY REQUIREMENTS STUDY AIRLIFT MODELS

In 1991, the Joint Staff evaluated U.S. strategic lift to determine whether it was adequate to deploy U.S. forces in time to achieve national military objectives. The abstract models the Joint Staff used for comparing alternative aircraft fleets for strategic airlift had simple criteria and algorithms. Their criteria included gross weight carried, airspeed, mechanical reliability, and range; it did not include outsize or oversize cargo.[5] Either of these types of cargo causes the U.S. Transportation Command to use aircraft contrary to the model, i.e., the transport aircraft cannot load cargo to the study capacity. Thus, the models did not accurately predict the actual behavior of the strategic airlift fleets under true operational conditions.

Despite their limitations, the models were still very useful for the Joint Staff. They enabled them to determine the relative differences in performance among the aircraft fleet alternatives. The Joint Staff did not mistake these insights into relative performance differences among fleets for absolute outcomes. While they could conclude one airlift fleet had 30% more capacity than another did, they knew they could not say that the first fleet would deliver X tons of supplies in seven days while another took ten. Although we would like the model to predict faithfully what will happen, we can often settle for models that show differences in relative performance, despite their inability to evaluate absolute performance.

Models differ in their ability to predict what will happen in the real world. Some models do not predict the absolute outcome of events very well, but they are still useful as long as they display a relative difference in performance among the alternatives that will carry into the real world.

Types of Analytic Models

DoD uses many standard models for analysis. For example, Joint Simulation System (JSIMS) provides a validated computer-simulated environment for use by the CINCs, their components, other joint organizations, and the Services to jointly educate, train, develop doctrine and tactics, formulate and assess operational plans, assess war-fighting situations, define operational requirements, and provide operational input to the acquisition process. Another example is actually a suite for four simulation models, JQUAD, which contains electronic warfare, command and control, network, and operational intelligence models. These models, along with numerous others that have been validated by the Pentagon, establish methods for the most frequent analyses by using common frames of reference. Using an already-accepted model automatically focuses discussion on the unique aspects of the decision whereas with a new model, we will have to gain acceptance before we can advocate our preferred alternative. Therefore, we should always consider modifying existing models to fit our decision rather than building a new model from scratch.

Below we list some of the more common types of analytic models that can be used for defense resource allocation decisions. Which model we select depends entirely upon the situation; an appropriate fit between model and problem is paramount. Because models vary in abstraction, com-

5. Outsize cargo, e.g., tanks, exceeds 9.75 feet in width, 8.75 feet in height, or 90.8 feet in length; it is the largest class of air cargo and it fits into C-5 and C-17 aircraft but not C-141s. Oversize cargo is typically a single item, like a pickup truck, that does not exceed the size of a standard 463L pallet but does not allow the air-craft to stow cargo to its maximum capacity or efficiency. C-141s can carry oversize cargo. Source: Military Airlift: Airlift Planning Factors, AFP 76-2 (C-1), 1982, p. 4-5.

plexity, and their ability to predict in different situations, we must have a clear problem definition and a thorough understanding of how our criteria interact before we select a model.

We often run models using scenarios as backdrops. Scenarios are situations, a collection of boundaries, including facts and assumptions from the Definition Phase, and other necessary conditions for running the model, such as location, time frame, sample size, etc. We may specify scenarios for the problem we are solving or have the analysts develop them based upon existing or predicted scenarios, e.g., the Defense Planning Guidance includes two appendices of illustrative scenarios (one current and one future) for force structure planning; the Combatant Commanders test their concepts of operations in scenarios loaded into large models.

Again, the names of these models are less important than understanding their character and understanding how we can apply them to different types of decisions. We also present them here because analysts often use this terminology in their descriptions and proposals.

DETERMINISTIC MODELS

These models require a thorough understanding of causes and effects in the environment or problem we are modeling. We change one or two key input variables, leave the other variables stable, and produce an outcome resulting from the input changes: input ***a*** leads to output ***b***. We use deterministic models when accurate prediction is especially important and we have a high level of certainty about the controlled variables in the model.

Many simulators use deterministic models. In an aircraft flight simulator, moving a control in a particular manner causes change in the flight characteristics related to it. The model determines the overall effect the control adjustment will have, and reacts accordingly. Deterministic models, assuming they are built correctly, are very reliable predictors—they will produce the same result under the same circumstances every time. Therefore, we must decide if that is also true of the portion of the real world we are trying to describe before we commit to a deterministic model.

INVENTORY MODELS

Used primarily by logisticians to manage stock levels, these models play an important role in force planning, particularly in procurement, because life cycle costs are dramatically affected by spare parts and energy consumption: their cost, usage rate, storage, and delivery. To be effective, these models require solid estimates about user consumption. Generally, inventory models contain two or more competing cost curves, e.g., storage cost and transaction cost for spare parts.

Using a naval example, storage cost is the expense of maintaining an inventory of spare parts for rapid issue to the Fleet. Transaction cost is the cost of obtaining an item directly from a supplier on demand; generally this takes longer than an internal transaction within the Navy and is more expensive because there are no price breaks for large volume purchases. But if we store too many of these spare parts, we have several problems. First, the Navy may have too much purchasing power tied up in inventory—stocking the inventory imposes opportunity costs in other areas. Second, warehousing them creates costs by itself. Finally, if these parts are technologically perishable, we will waste resources if they are never consumed and they have little disposal value. The analyst seeks to find the lowest cost over the life cycle of the system to balance the two costs and recommend an inventory level to the Navy that optimizes responsiveness (adequate inventory within the Navy) and transaction costs (frequency of replenishment of that inventory).

ALLOCATION MODELS

Allocation models examine the most efficient assignment of resources to tasks. Typically, we use spreadsheet programs to explore the effects of a change in one area upon another. In DoD, we use allocation models to solve assignment problems wherein we have a number of tasks and a number of units that can fulfill them. When a CINC provides guidance for a quarterly schedule for his ships, he considers current operational requirements, exercise participation, in port maintenance requirements, and port visits in support of his Theater Engagement Plan. These requirements and the list of ships available could be built into an allocation model that would optimize scheduling, or at least provide a rough schedule to use as a starting point. Allocation models are also useful for solving transportation or network problems in which the analyst seeks the most efficient path from a starting point to an outcome. The variables in the model behave very much like the values in utility curves (see Chapter 6).

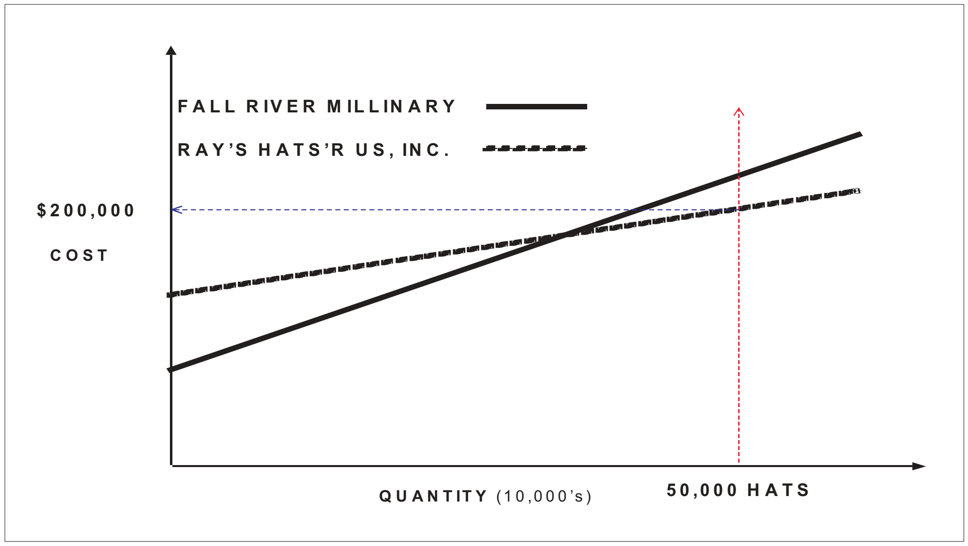


Figure 7-2. Allocation Model.

An example of an allocation model is shown in figure 7-2. In this case, we are buying headgear to stock a uniform store and we have two competitors that have provided price curves. We need 50,000 hats (fixed effectiveness), so we select the most optimal solution by reading up and then across to identify the lowest cost—Ray's at $200,000.

STOCHASTIC MODELS

Stochastic models are always dynamic or interactive; they incorporate time, randomness, and probability theory. They are very useful when we have high degrees of uncertainty, when input ***a*** yields output ***b, c, d,*** or ***e***—or any combination of them. One branch of stochastic models involves queuing processes. Queuing models derive their name from their initial applications in the service sector, i.e., they were used to identify the number of passenger gates and their arrangement in airports. To build this kind of queuing model, the analyst first in-puts the problem boundaries: the service or process time (fixed and known) and the behavior rules for processing the people in line, including decision rules such as: First In, First Out; Last In, First Out; or Very Important People To The Head Of The Line. Then the analyst designates the number of service stations (the range of solutions) for different runs of the model. The model uses stochastic methods to input customer arrival times (the random or probabilistic event) with a variety of surges and slack periods (random or designed by the flight schedule) as the model runs. The output of each run is information about customer waiting times: average, longest, mean, etc. An airline using this model could set a goal for an average waiting time and then use the model to predict how many customer stations it needs manned to satisfy loading at different times for different days.

In DoD, queuing models help us plan the overall capacity we require for maintenance and support of a force structure. To support the 1997 Quadrennial Defense Review, DoD constructed a wargame called Dynamic Commitment to examine the demands that might be placed against the U.S. military in the next fifteen years. The analysts constructed a series of scenarios from major theater wars to a variety of peace operations and let the model generate their sequencing stochastically in accordance with some rules, i.e., no more than two major theaters wars simultaneously and at least five years between major theater wars in the same theater. None of the strings was meant to be a literal prediction of the future; rather the analysts used the results of their many runs to identify the character of the force structure that was most likely to be successful at meeting every requirement. (Unfortunately, the force structures identified exceeded the resources likely to be available by a wide margin.)

Markov chains are another stochastic modeling tool. Markov chains exist when the probability of one event happening depends on what happened in the event that immediately preceded it. These are the mathematical equivalents of the branches and sequels we use in operational planning. For example, service-recruiting targets are a function of force structure requirements and retention; changing either will affect recruiting goals. Stochastic modeling has become prevalent with the use of computers that can manipulate a plethora of data, equations, alternatives, events, and possible outcomes; therefore we use these models to support wargaming.

COST ANALYTIC MODELS

Cost models range from the very simple to the extraordinarily detailed. Some use advanced mathematical techniques, others only basic arithmetic. Some require extensive computer support, others analysts build manually or with simple spreadsheets. Remember that cost estimating methods tend to overlook costs that cannot be measured in dollars and these other types of cost are often more important to us than dollars alone.

For existing weapons and support systems, we can estimate cost using historical data. However, for many force-planning decisions, the systems do not yet exist. Fortunately, there are numerous cost estimating methods that can be used to predict future costs. Three of the most common are the analogy, parametric, and industrial engineering methods.

Analogy Method

When detailed cost data is not available, an analyst may estimate cost by making direct comparisons with similar existing systems. For example, using the analogy method, we can approximate the value of surplus land on a DoD installation based on the sales of similar property nearby. We often estimate low-value equipment proposals, commodity purchases, and operating and support expenses using analogies. This method is also very effective for estimating the cost of off-the-shelf equipment where comparable prices are as close as the nearest catalog. In order to use the analogy method for new or complex concepts, an analyst needs considerable expertise and judgment. The less compatible the subject and the model, and the older the existing comparator, the less confidence we have in this kind of cost estimate.

Parametric Method

We may deem it impossible to find an appropriate analogy to use to estimate cost for a new item. However, we may be able to identify characteristics or parameters of the new system that are similar to the characteristics of other existing systems. Using those carefully identified parameters, we seek a cost estimating relationship that we can project onto the new acquisition. The cost estimating relationship sets this method apart from the analogy method. It is a mathematical expression that relates one or more particular acquisition characteristics to cost, e.g., cost per ton for the construction of a ship. Note that the cost-estimating ratio itself may be based on an analogy; we may estimate the cost of a new government warehouse, larger than any previous building we have contracted, by multiplying the area times the cost per square foot of an airplane hangar or large civilian warehouse.

We use the parametric method in DoD for estimates early in the Defense Acquisition System. Parametric estimates can be very accurate when they are based on realistic, historical experience, as demonstrated in the accuracy of F/A-18E/F cost estimates, which were based on the costs of the C/D model. Moreover, we can calculate the cost estimate quickly once we establish the cost estimating relationship. Parametric costing may result in pessimistic estimates if we do not adjust the formulas based on historical experience for improved production methods or recent lessons learned.

Industrial Engineering Method

The industrial engineering cost estimating method is often referred to as the bottom up approach. An analyst using this method consolidates estimates for various segments of a project into a total estimate for the entire project. Government analysts estimating the cost of a new building use this method by estimating the structural, electrical, plumbing, heating and air conditioning, and other component costs of the projects. They may break each of these estimates down further into sub-components such as labor, materials, equipment, etc. The industrial engineering method is the most thorough way of estimating cost, but it can be quite time consuming.

Evaluating the Model

Before the analyst runs the model and we compare alternatives, we will evaluate the model to ensure it reflects how we think the criteria behave and interact. First, we review the Definition Phase to ensure the guidance we gave the analyst conforms to our analytic objective and that our analytic objective still makes sense. Then we review the analyst's model proposal to ensure it aligns well to the analytic objective, e.g., we do not want to use a complex stochastic model to evaluate a simple decision about bulk commodity purchases. This kind of mismatch happens most often when we use an existing model for a new decision situation. Then we evaluate the model's level of abstraction, complexity, and predictive qualities in terms of validity, reliability, and practicality. When we are satisfied with the qualities of the model, we should obtain the decision maker's approval before proceeding further.

MODEL VALIDITY

As we examine the validity of our model, we ask whether it captures the most important behaviors of the alternatives at the right level of resolution—does it model the right things? Do the criteria reflect our perceptions of reality? In a weighted model, do our utility curves and weights reflect our values? The boundaries in the model must be consistent with the elements we identified in the Definition Phase. It must model the alternatives objectively. We must understand the predictive qualities of our model to ensure it helps us distinguish among the outcomes and we must have confidence that the models' projections are consistent with the real world. Finally, the model's level of complexity must be appropriate for the decision maker.

We need to view the model as a totality, also. We can get mesmerized by the detailed evaluation of criteria to a point where we lose sight of the analytic objective. Air campaign planners, used to trading off strengths and weaknesses of tactical aircraft, sometimes need to be reminded to use models that are robust enough to include tactical missiles, bombers, and attack helicopters.

MODEL RELIABILITY

Where reliability is concerned, we are interested in the model's behavior: does it measure well? The internal consistency of our model determines whether we are confident that the results of the model (predictions) will be the same whenever the model is used under similar circumstances. We must be able to measure the criteria well at the model's level of abstraction.

The model may be affected by measurement errors as we collect data on the criteria, especially if we fail to measure with enough resolution. We must determine how much error is tolerable while running the model and whether we must measure some criteria more precisely than others. If others are providing data, we must be satisfied with its accuracy. Generally, the more abstract the model, the simpler it becomes and the more forgiving it is of our measurement errors.

MODEL PRACTICALITY

Some models are more costly than others, and we seek to balance realism (validity and reliability) with cost as we address the model's practicality. Reducing cost to avoid the difficulty and expense of real world testing is our reason for modeling in the first place. The resources we consume in modeling should be commensurate with the importance and urgency of the problem to our organization.

The more abstraction we accept (the further we move away from reality) in the model, the more vulnerable we are to criticisms that the model does not reflect the real world. In addition, our results are more difficult to "prove." If the model's predictions are too unreliable, we will have to improve its data, reduce its level of abstraction or make the model more complex. We add complexity most often by making the algorithm more intricate, by adding variables (not necessarily criteria), or by increasing the level of detail in their measurement. All of this takes resources—time and money.

Analytic Models and the Information Age

The ability to store and retrieve data electronically from sources all over the world has greatly improved the quality of analysis in general and models in particular. Their validity and reliability are increasing as computers allow increasing complexity without degrading reliability significantly and at a reasonable cost. But computers may also conceal errors if we fail to understand the assumptions made by programmers and how they related our criteria to one another. “Garbage in, garbage out,” requires we be able to identify what is garbage.

DECISION SUPPORT SYSTEMS

Decision support systems are interactive software we run on computer hardware ranging from mainframes to networks to laptop personal computers. Decision support systems are very useful for organizing and manipulating subjective inputs from multiple participants in a decision and converting them into preferences for alternatives. The simpler systems help us build weighted models to compare procurement alternatives; the more complex decision support systems help us make force structure and policy decisions.

Decision support systems allow us to introduce structure and rigor to very complex problems and they are especially valuable when we cannot adapt other techniques to model the problem. For example, the Decision Strategies Department of the Naval War College, which has professional facilitators and uses a network of laptop computers, has examined policy alternatives for issues like confidence-building measures between Greece and Turkey and NATO enlargement. Because of the constant requirement for subjective judgments, we strongly desire the decision maker to be present when we use a decision support system for a policy issue.

NETWORKED MODELS

Our ability to exchange data through computer networks makes data commonality feasible and the process of data collection much easier. Decision makers in different locations can view the same spreadsheets and do sensitivity analysis during a teleconference from their workstations. The Joint Strike Fighter program is using this shared data base capability. The contractors and the program office use a common cost model; the DoD Program Manager can discuss cost data with a contractor while they both view the same database, tremendously simplifying coordination and reporting.

Using the Model to Evaluate Alternatives

Once we are satisfied with the model, we insert the alternatives and evaluate each. Recall that sometimes we have the alternatives before we build the model. In this case, we may have tailored our model to highlight the differences between the known range of alternatives and our foreknowledge may affect our criteria selection in particular. Because we use criteria to discriminate among options, we are unlikely to select an attribute whose value is equivalent for each alternative as a criterion. As we run the model, however, new alternatives may emerge and that may require us to re-evaluate our criteria and adjust the model.

When we create or learn of the alternatives after we build the model, the application is more straightforward. Sometimes, however, an unusual alternative arises after we have assembled our model that forces us to reexamine it, either to add new criteria or to identify a new requirement we need to apply to all the alternatives. The new criteria may not have discriminated among the previous options because they scored similarly. The new requirement may be necessary to exclude impractical solutions, e.g., a training range may be ideal in every regard except it is too far from homeports.

After the model runs, we have its results. Depending upon the nature of the problem and the model we used, they may vary from identifying a single preferred option to a hierarchy of scores for different alternatives, or a series of tables. In any event, we should be able to interpret them easily and explain them to others with clarity as we did with the radio example in the previous chapter and which we will continue below. We should not hesitate to stop and examine the model if its results defy easy explanation. While the possibility exists of new and exciting insights, it is more likely we have made a mistake and we need to find it and correct it.

When we are satisfied with the results, we need to create reports and briefings to support the decision maker. The seniority of the decision maker, the time available for briefing, and the magnitude, urgency and importance of the problem we identified in the Definition Phase will determine the amount of detail we present. Naturally, we should be able to explain the connective tissue from the most general of slides down to the measurement data if need be, just as an Executive Summary derives from a formal report and the report is based on modeling and data (often included in appendices).

Sensitivity Analysis

After the analysts run the model and results emerge, we often observe that some facts, assumptions, or criteria have an unusually strong influence on the outcome. Also, the analysts may not have data for some variables when they run the model, so they assign them arbitrary values, effectively making their own assumptions. We need to know how sensitive the results of the analysis are to changes in the values of variables, particularly if those values were estimated. To establish how changes in the value of a particular variable affect outcomes, we fix the values of all the variables in the model except the one under study. We then run the model several times, using a different value for the variable under study—high, low, and medium values for example—to see how changes in that variable affect the results. This process is called sensitivity analysis.

We may use sensitivity analysis in many ways during the Analysis Phase. First, we may change the boundaries of the problem or the initial conditions by altering facts or assumptions. For example, during Dynamic Commitment, changing the scenario queue to allow only one major theater war at a time results in a significantly smaller force structure set. We may also directly change the weights in a weighted model or the values of a criterion for different alternatives to explore variants and combinations of options. We can use sensitivity analysis to examine a criterion through the estimated range of its measurement error to see if we need better data.

Computers enable us to conduct a vast amount of sensitivity analysis rapidly and easily. We can vary almost any data or assumption in the model to determine whether changes are important to the results. In addition to its information value, sensitivity analysis is a powerful cost saving technique. For example, one of the variables in the model may be very difficult and expensive to measure. If we establish a range of probable values, run the model, and the preference between the alternatives does not change for these different values, the model is insensitive to that variable and we can use an assumed value without undermining the analysis.

If the outcome does change with different values, it is sensitive to that variable and we need to find a way to measure it directly or through a proxy. If the sensitive variable is an assumption, our last resort may be to display multiple sets of results for the different values of the assumption. For example, if we are comparing the life cycle costs of aircraft alternatives, each with different fuel consumption rates, the relative difference among options may be sensitive to our assumed price of jet fuel. We can check for sensitivity by running the model with our lowest estimated fuel cost and again with our highest estimated fuel cost to see whether the cost rankings of the alternatives change.

Another way to employ sensitivity analysis is to change the weights in a weighted model (without changing the values or scores of any alternative's criteria) to see if a change in weight alone changes the relative rank order preference of the alternatives. For example, let us return to last chapter's portable radio scenario and the output of its weighted model (see figure 7-3 on next page). We reproduced the alternatives and criteria, with their weights added in parentheses, in table 7-1.

RADIO	PURCHASE COST (25)	TOC (15)	RANGE (15)	WEIGHT (20)	SECURITY (15)	RELIABILITY (10)	TOTAL (100)
POPIEL 1995	23	15	0	20	0	0	58
WHAMMO 3000	11	7	12	16	15	6	67
ZONKER 101	4	14	15	6	15	10	64

Table 7-1. Radio Alternatives.

Recall that when we ran our weighted model, the Whammo 3000 scored the highest. The Popiel 1995, a lightweight inexpensive option with the minimum acceptable performance, scored poorest. What happens to the results of using the model if we change the weights to reward cost and effectiveness equally, i.e., weight each 50 vice the original 40 and 60, respectively? Assuming the weight changes are spread proportionally down to the criteria in the lowest tier and the performance of the alternatives does not change, we get a new preference for the Popiel 1995 as shown in table 7-2:

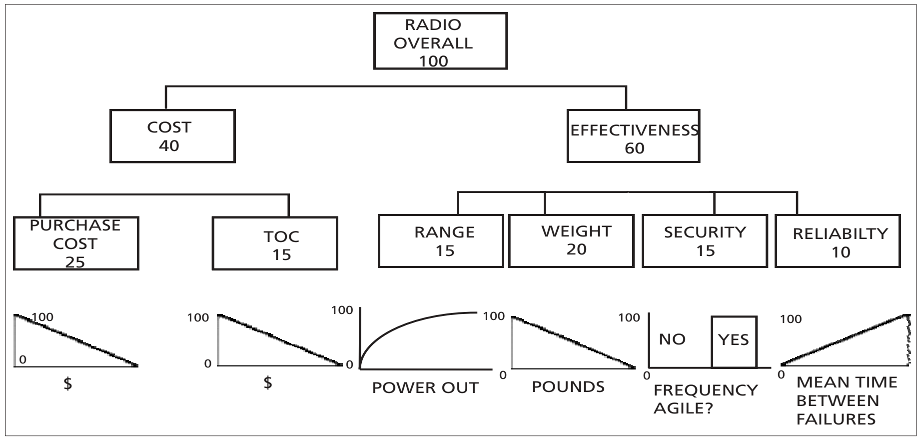


Figure 7-3. Weighted Model for a Portable Radio

		INITIAL COMPARISON			ADJUSTED COMPARISON	
RADIO	**COST (40)**	**EFFECTIVENESS (60)**	**TOTAL (100)**	**COST (50)**	**EFFECTIVENESS (50)**	**TOTAL (100)**
POPIEL 1995	38	20	58	47.5	16.7	64.2
WHAMMO 3000	18	49	67	22.5	40.8	63.3
ZONKER 101	18	46	64	22.5	38.3	60.8

Table 7-2. Radio Model Sensitivity to Cost and Effectiveness Weights.

Thus, we see the importance of choosing our weights carefully and rationally. When we see how fairly small changes in weighting can lead to large changes in outcomes like the shift in model-preferred alternatives between a high-cost, high capability radio to an inexpensive, less capable radio, we must also ask about the validity of the weights in the model. Which is really more important to us—cost or effectiveness? We can also make a strong argument that by building the model before we know the alternatives we are more likely to reflect our organization's values impartially. Further, we can understand why, if we use someone else's model, we need to understand how it works before we accept its results.

We can also use sensitivity analysis to see how much change is necessary in one variable of one alternative to make it the preferred choice—or determine that no amount of change in that area will make it so. Returning to the hypothetical radio scenario (with the original weights in figure 7-3), consider the Zonker Company's situation: it is very competitive with the Whammo model. What can it do to overtake Whammo within the model? The Zonker 101 has achieved maximum performance in three of the four effectiveness criteria, but it scores poorly under Weight. If Zonker can improve performance in this area by lightening a calculable (if they know the shape of the utility curve for weight) number of pounds from the radio, they can achieve a higher score than Whammo. Likewise, they may be able to reduce their profit margin in order to decrease their selling price and become more competitive.

CASE STUDY: THE ANALYSIS PHASE—MODELING USMC MEDIUM-LIFT REQUIREMENTS: THE V-22 OSPREY AND HELICOPTERS

Returning to the V-22 and helicopter analysis, we now examine how the Institute for Defense Analyses evaluated their analytic objective: to compare the V-22 and helicopter alternatives on the basis of cost and operational effectiveness. IDA used a cost-risk-effectiveness approach for their modeling. IDA studied cost separately, and then combined cost and effectiveness to achieve the analytic objective. What little risk they examined was built into their effectiveness measurements. IDA's analytic method is summarized in the diagram below:[6]

Recall that IDA fixed cost by creating two sets of equally expensive aircraft fleets to compare the V-22's and helicopters' effectiveness. The first set was sized to the Marines' desired fleet of 502 V-22s and the second to what DoD was willing to budget for medium lift—356 V-22 cost equivalents.

IDA evaluated the effectiveness of the aircraft fleets in the missions mandated by Congress and in an additional area, Anti-Submarine Warfare, specified to them by DoD. The missions were:

- Amphibious Assault
- Sustained Operations for Logistics Support
- Hostage Rescue/Raid
- Overseas Aircraft Deployment
- Combat Search and Rescue
- Special Operations
- Drug Interdiction
- Anti-Submarine Warfare

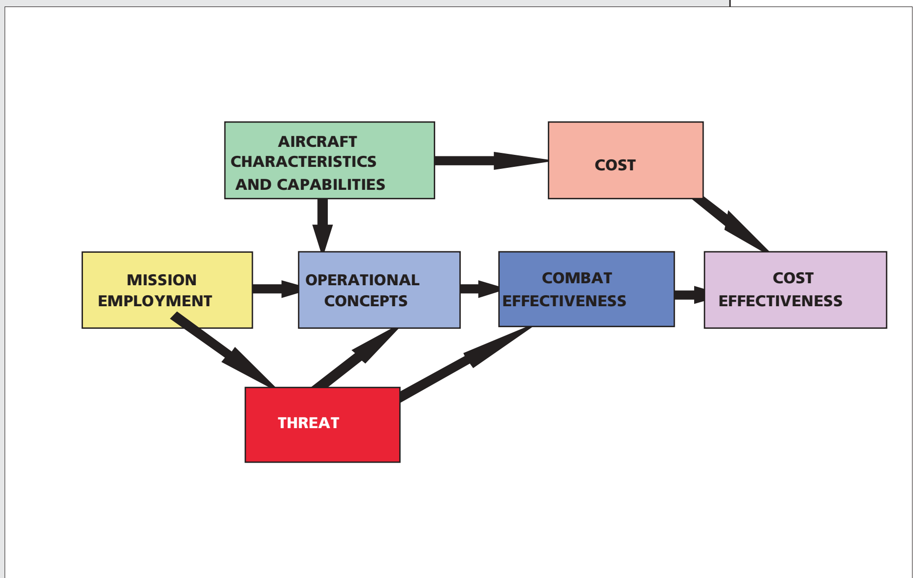
Because the majority of medium-lift aircraft are intended for the amphibious assault role, IDA accorded it particular attention. IDA evaluated the air defense threat in each mission scenario and developed operational concepts that they coordinated with the services and the Joint Staff to ensure they were modeling aircraft employment realistically. Using their abstract operational concepts, IDA estimated the performance of each type of aircraft—the V-22 and the six helicopter options—to determine combat effectiveness. They ran a very large set of excursions to study the fleets' performances in the scenarios.

The Institute for Defense Analyses used at least one model in each of the eight missions. We will concentrate on the amphibious assault scenario because it is the most important to the overall analysis and because IDA used the most complex models for that mission.

IDA used an existing deterministic model to analyze amphibious assaults. This engine was the Amphibious Warfare Model, a 1970's era computer simulation of a conventional theater assault, developed and updated continually by the Center for Naval Analyses. To examine the performance of the options under varying conditions, IDA selected two Department of the Navy case studies and built two corresponding vertical assault forces, each attacking under different battlefield conditions. The assault forces began on amphibious ships 50 nautical miles from the landing zones in both scenarios. A notional Third-World Soviet-style Motorized Rifle Division opposed the Marines in each.

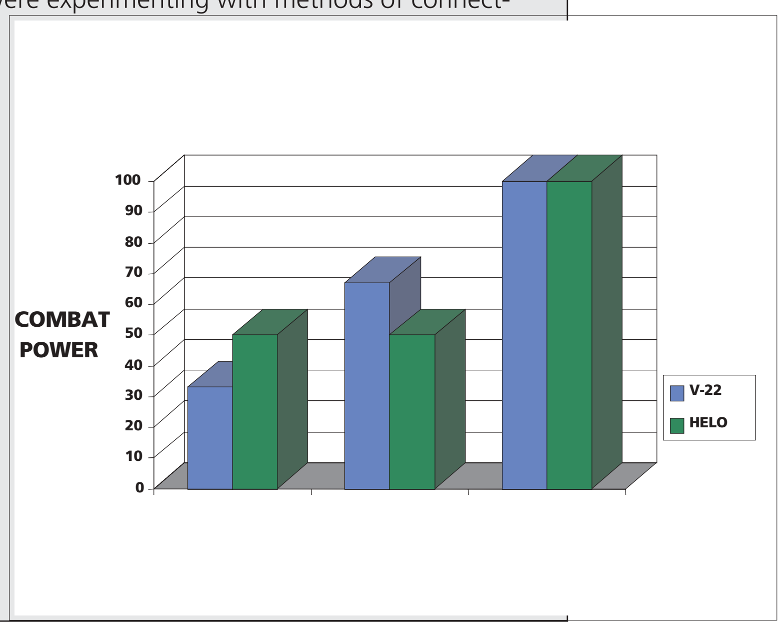
6. Simmons, L.D. Et al, Assessments of Alternatives for the V-22 Assault Aircraft Program, Executive Overview, Institute for Defense Analysis, 1991, pp. 3–4.

IDA made some critical assumptions before running their model. One of the most controversial was a change to Marine Corps doctrine. In 1990, Marine planners assumed the vertical assault would land in only two waves to ensure the first wave had enough combat power to survive enemy reactions until the remainder of the force landed. With the V-22 lower option of 356 aircraft, the Marines could not land 50% of the vertical assault force in the initial wave. IDA assumed that the Marines would accept delivery of the vertical assault combat power in three vice two lifts if the build-up time was not compromised. They reasoned if the Marines' desire was to get a given capability ashore within a time span from H-Hour to time T, the V-22 (with its superior speed) could deliver the same force in the same time frame in three lifts, vice two for the helicopters, and still meet the Marines' requirement, as shown below.



Some critics challenged this assumption, notably Dr. David Chu, Assistant Secretary of Defense for Programs Analysis and Evaluation (see Appendix 3). He noted in his congressional testimony that the 356 aircraft V-22 fleet would have to generate a historically high sortie rate from the assault ships to achieve the build-up in combat power in the time the scenario required. The Marines, however, concurred with IDA's interpretation and skep-tics accused them of redefining doctrine to suit a procurement goal. DoD ultimately decided it was up to the Marines to define their doctrine and if they chose to modify it, that was an internal Marine Corps decision.

A second major assumption the Institute for Defense Analyses made concerned the method the CH-53E heavy lift helicopters, present in each helicopter fleet option, used for lifting external loads slung underneath the aircraft. The Marines were experimenting with methods of connecting two vehicles together as a single, stable load beneath the helicopter to reduce the number of sorties needed during an assault. If they were successful, they would reduce the number of V-22 (and medium helicopter) sorties dedicated to lifting vehicles. At the time of the IDA study, the Marines had not tested these methods at sea. Skeptics were concerned that linking the vehicles would be impractical on darkened, rolling ships and that unlinking them in a landing zone under fire would be too hazardous.



IDA, as in the 1990 Navy study, assumed CH-53Es delivered half the vehicles in dual lifts for smaller assault forces. They assumed all the vehicles would be in dual slings to lift the larger assault forces. IDA did sensitivity analysis to see how the number of vehicles in double slings affected the results of their model; they found the more vehicles that were double-lifted by the CH-53Es, the smaller were the delivery performance differences between medium-lift options, but that the rankings remained the same.

Returning to the assault scenarios, the first study situation was Department of the Navy (DoN) Lift I[7] from 1983. It had the Marines make a night assault in rolling terrain. The low aircraft flight profiles reduced the effectiveness of the enemy air defenses because they were masked by terrain. It also assumed poor reaction times by the defenders. The second scenario was from a 1990 study, DoN Lift II[8]; it set the assault force against a faster-reacting defender about two thirds as well armed as its 1983 counterpart. This time the assault happened in daylight over flat terrain with better fields of fire for the air defenders. Not surprisingly, aircraft casualties were higher in the second case and the casualty differences between aircraft options were larger. IDA ran 388 excursions, varying assault force compositions, tactical factors, threat, and terrain for each aircraft fleet. IDA measured the percentage of the Marine vertical assault element lost attaining a 3:1 advantage in combat power over the defenders to compare and rank the medium-lift options.[9]

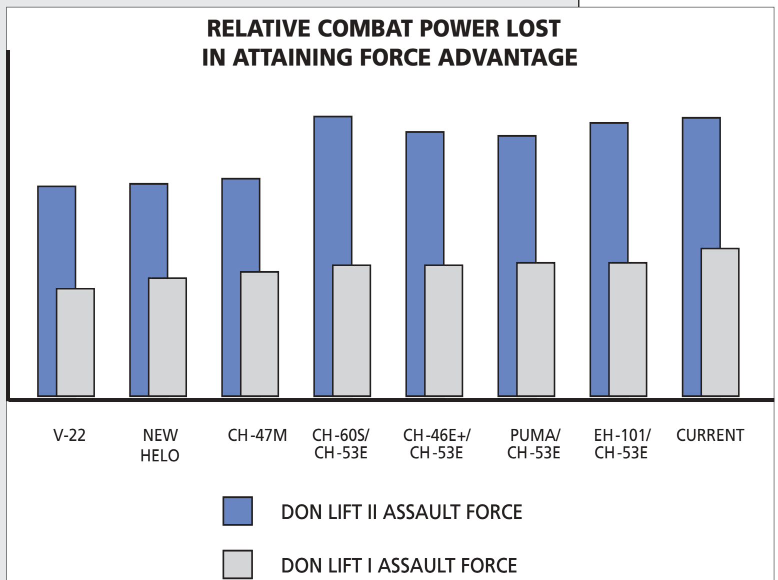
Using both assumptions, IDA ran the model for the aircraft fleets in the two scenarios. With survivability as the principal measure of effectiveness, the V-22 outperformed the helicopters in the amphibious assault mission. They displayed the results in a series of bar graphs, one set for each fleet in each assault case, as shown below (figure 3 from IDA's Executive Summary).

These bar graphs represent the results of the 388 combinations of enemy force composition, tactical factors, threat, and terrain that IDA explored. Those results all fell between the ranges of these bar graphs. In the Amphibious Warfare Model, the size, speed, design, and length of time an aircraft was exposed to enemy air defenses during each possible engagement determined its casualty rate. The V-22, with its higher speed, moved through air defense engagement envelopes faster than the helicopters, therefore it took fewer casualties (although the smaller, harder-to-hit helicopters approached the V-22's survivability rate). Moreover, if DoD opted for the smaller helicopter fleets, they would also have to buy 200 to 260 large, more vulnerable CH-53E helicopters to compensate for the limited external load capability of the smaller helicopters.

Next we will evaluate the validity of the Amphibious Warfare Model for assessing helicopters and the V-22—is this the right model for comparing the medium-lift aircraft alternatives? The level of abstraction of the model for this application is very high because IDA used a very small portion of a very large model for this study. This portion distilled the effectiveness of the aircraft options into a single MOE, survivability, and used a very simple combat engine to evaluate each aircraft. This forces us to ask whether size, speed, length of time in the air defense envelope and the resilience of each aircraft to withstand battle damage are the only important determinants of aircraft effectiveness. How will the V-22 interact with other Marine aircraft for flight operations (flight deck crew turn around time) and for long-range assault (since it can outrun its attack helicopter escort)? Is the number of deck spots important to generate sortie rates? Is unit integrity of the passengers or unloading time important in the landing zone?

7. Department of the Navy Long Term Amphibious Lift Requirement and Optimum Ship Mix Study, Office of the Chief of Naval Operations/Headquarters Marine Corps, 25 May 1983, CONFIDENTIAL.
8. Department of the Navy Integrated Amphibious Operations and USMC Air Support Requirements Study, Office of the Chief of Naval Operations/Headquarters Marine Corps, 5 April 1990, SECRET.
9. IDA ran additional iterations to examine 2.5:1 and 3.5:1 build-ups; the preference rankings of the alternatives remained the same, i.e., the model was not sensitive to how much combat superiority the Marines required.

The behavior of the air defense forces is also greatly simplified; it is based on the air defenders' reaction time. The deployment of the air defenders was arbitrary, but it was the same for all the assault aircraft options in the model. These are all clearly shortcomings in replicating reality; but are they fatal? This is where we must insert professional military judgment to evaluate validity and determine if the relative performance differences of the model carry into the real world. In our opinion (and DoD's, including the Marine Corps), the level of abstraction is appropriate for this 1990 decision and the simplifications were acceptable.



The model is predictive and it needs to be; it is forecasting aircraft casualties during the build up of the assault force in the two cases described above. How accurately we think it predicts depends on our confidence in the assumptions we discussed earlier and our acceptance of this high level of abstraction. We think it will predict the relative behavior of the aircraft alternatives accurately. IDA could make this model more complex—it could incorporate flight operations variables, a more complicated combat engine and more types of air defense weapons. But would these improvements change the outcome of the model output? Probably not.

Now we turn to the reliability of this model—does it model accurately and consistently? We have an inherent reliability problem whenever we rely on contractor projections about aircraft that have not been built yet, in this case the V-22 and the new helicopter. The values for the variables in this model were readily available to IDA from existing databases or were provided directly by the contractors and we have a high level of confidence they reflect real world performance. IDA ran the model hundreds of times and the outcomes were consistent throughout the study. The overall reliability of this model was very high.

This was an important and urgent study; practicality was central to many of IDA's decisions about the model. They knew the six previous studies comparing the V-22 and helicopters had not provided enough information to finally decide this aircraft selection; they felt compelled to add new knowledge to support the decision makers. IDA needed to conserve resources, especially time, producing this analysis. They cleverly adapted existing studies and an existing model to compare the aircraft options, tools previously accepted by the major decision participants. Discussion and controversy quickly focused on the limited number of assumptions and the results of using the model, which was what the participants desired, i.e., they were not distracted examining and debating the model. Enhancing the model to reflect reality in more detail, as described above, was not worthwhile because even if IDA increased the level of detail it would not change the rankings of the aircraft options. The Institute of Defense Analyses scored well in practicality with this study.

IDA presented their findings in six volumes, including the Executive Overview. They presented most of the results in tables and graphs and displayed the utility of the different options, arranged by fleet cost and alternatives. IDA briefed the services, Joint Staff, and Defense Secretariat of their results and eventually testified before Congress.

Throughout the Analysis Phase, IDA verified they were executing the decision maker's desires. An Office of the Secretary of Defense Steering Committee held five meetings during the course of the study to validate IDA's plan and monitor its progress. Importantly, IDA validated their scenarios with DoD's subject-area experts to include military judgment.

Summary

As our decision becomes more complex, analytic models become less capable of providing clear-cut, definitive answers about how we should choose among alternatives. But even then, well-constructed analytic models provide important insights about how and why the alternatives perform as they do. Combined with professional judgment, this kind of information can guide our choice of courses of action.

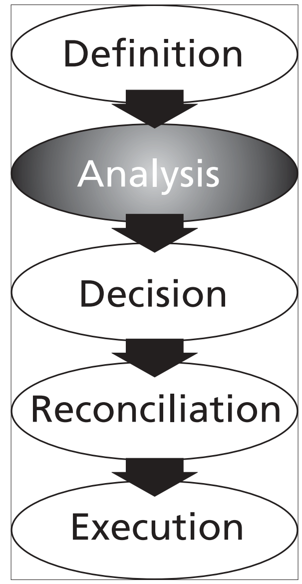
Models are important tools that facilitate our decision making by simplifying complex problems, making them easier to understand, change, and manipulate. By using models, we reduce the cost and effort of evaluating alternatives by substituting modified or imaginary environments for actual conditions. Based on the nature of our decision, we select the type analysis we are going to use: exploratory analysis and concept studies for new ideas, cost-risk-effectiveness models for analysis of alternatives, and causal analysis for policy options.

We select or build our models on the basis of the decision we are making, the type of analysis we are doing, and our required levels of abstraction, prediction, and complexity. We prefer to use existing models rather than creating new ones, but we will not force a fit. As with the criteria, we evaluate our models on the basis of validity, reliability, and practicality. We perform sensitivity analysis to identify which variables have the greatest effect on the results of comparing alternatives, enabling us to target changes to options (or the model) to have the greatest effect.

For all their strengths, good models do not guarantee we will make good decisions. Models can have significant shortcomings, especially if they are used incorrectly. Choosing or building the right model to use in a particular decision situation is highly dependent upon the judgment, experience, and collaboration of the decision maker, action officers, and analysts.

CHAPTER 8

FORCE-ON-FORCE ANALYSIS



With many calculations, one can win; with few one cannot. How much less chance of victory has one who makes none at all!

—Sun Tzu, 400-381 BC, *The Art of War*

FORCE-ON-FORCE ANALYSIS INVOLVES determining how effective a military force is likely to be in combat situations, what factors are most important in determining that effectiveness, and how changes in the force, the adversary, or the combat situation change the likely outcome of combat. Force-on-force analysis is the heart of force structure planning, one objective of which is to develop forces that will prevail in combat. Thus combat effectiveness, expressed in many different ways, is the fundamental criterion we use to compare competing weapons systems, doctrines and operational concepts, force structures, theater Operational Plans (OPLANs), and military strategies.

Force-on-force analysis is the method the U.S. defense community uses to measure combat effectiveness short of committing forces to actual combat. Note that force-on-force analysis is not restricted to questions involving combatant forces only. All military functions, activities, capabilities, and organizations have the ultimate purpose of increasing the effectiveness of U.S. strategic and general purpose combat forces; therefore, we assess them also during force structure analysis in exactly the same way as combat forces themselves.

Formal and relatively abstract force-on-force analysis is a relatively recent invention. Militaries have used exercises in the field for analytical purposes only since the 19th century. The development of indoor force-on-force methods began in the late 19th century in Germany in the form of a board game called Kriegspiel (literally, "wargame"). Dr. Frederick W. Lanchester in Great Britain developed the mathematical roots of force-on-force analysis during and after World War I. In the U.S., formal, institutionalized force-on-force analysis began during the period between the World Wars. Here at the U.S. Naval War College, it took the form of elaborate wargames played at Sims Hall where naval officers developed and tested the amphibious doctrine and aircraft carrier tactics used to win the Pacific War. In the 1950s, mathematical force-on-force analysis using military operations research methods developed during World War II and the Lanchester equations became a basic tool for force planning. Nuclear weapons also lent themselves well to mathematical analysis. The computer has permitted enormously greater elaboration, detail, and speed in the mathematical models available for force-on-force analysis, but as we shall see, computers have not necessarily improved validity.

LANCHESTER'S EQUATIONS

One of the most famous attempts to predict performance of military forces with a dynamic mathematical model was attempted by Dr. Frederick W. Lanchester in 1916. Dr. Lanchester believed that the quantity and quality of military forces determine the outcome of battles and, therefore, that both must be included in any mathematical representation of combat. Briefly, Lanchester asks that we let

A = the numerical (quantitative) factor for force "A"

B = the numerical (quantitative) factor for force "B"

a = the qualitative factor for force "A"

b = the qualitative factor for force "B"

Lanchester postulated that two infantry forces of equal quality, but unequal in size, would inflict casualties on each other based on how many bullets they fired. The larger force would inflict more casualties on the smaller force with each volley; its strength and firepower advantage *relative* to its opponent's would grow after each exchange of fire. This strength advantage would grow as a function of the square of the quantities of soldiers on the two sides, e.g., a force of 2,000 soldiers is *four* times as powerful as a force of 1,000 soldiers. Eventually, the smaller force will be annihilated as it suffered more casualties than the larger force on each exchange while inflicting progressively fewer casualties on its enemy; in this case the force with 2,000 soldiers would take 268 casualties to eliminate the force of 1,000 soldiers. Lanchester named his equation the N-square law because of this attrition phenomenon.

Lanchester also considered how the quality of troops and other environmental factors affected attrition. He ultimately decided that attrition is proportional to the product of the square of the numerical factor multiplied by the qualitative factor of the other force, times a constant "K," which represents battlefield conditions such as weather, terrain, and the like for both "A" and "B." In other words, no single attribute included in K is more important than quantity with its exponential influence. Therefore,

$\mathbf{dA/dt = KbB^2}$ the change (decreased size) of A with respect to time describes the loss rate of force A, and

$\mathbf{dB/dt = KaA^2}$ describes the loss rate of force B

Note that the inclusion of "t" for time is what makes this equation a dynamic representation of combat. Much research has been devoted to testing whether Lanchester's equations and their implications have proven true in combat. In their simplest form, the answer is "No." This is not surprising since Lanchester developed his equation to represent combat as it existed immediately before and during World War I.

Lanchester's equation does a better job of predicting combat results when it is made more complex by adding additional terms representing morale, training, command and control, intelligence, and the like. Through a long process of adding and modifying terms, descendants of the Lanchester equations drive our current generation of force-on-force, campaign-level computer models.

Force-on-force analysis ranges from highly detailed engineering evaluations of individual weapons and their components to much more general assessments of global warfare. Between these poles lie such efforts as evaluations of aircraft, ships, and vehicles; analyses of the organization and effectiveness of tactical units and operational concepts; and assessments of the adequacy of joint theater forces. Force-on-force analysis is performed at every echelon of military decision making. A small unit commander planning an operation uses force-on-force rules of thumb to develop his plan. Similarly, a unified commander preparing theater OPLANs uses field and map exercises, wargames, and mathematical models to test alternatives involving different friendly and adversary forces, courses of action, adversary strategies, and other variables such as weather and terrain.

A military service staff in Washington preparing its annual Program Objective Memorandum analyzes its service's force structure to see if the existing and near-term programmed forces meet projected threats and the nation's obligations for forward deployments. Components of each military service use force structure analysis to develop and test alternative systems for acquisition to equip their operating forces. The Joint Staff performs continuous force-on-force analyses in the Joint Strategic Planning System that creates the nation's overall military direction. The multi-service, multi-CINC, and Joint Staff teams that together make Joint Warfighting Capabilities Assessments use force structure analysis to monitor the ability of U.S. forces and operational concepts to secure U.S. objectives and to warn when those capabilities are falling short. Additionally, the research arms of DoD, such as the service and national laboratories and the Defense Advanced Research Projects Agency, use highly technical force structure analyses focused on proposed weapon technology to determine how much leverage each technology might have on battle situations.

Theory of Combat

Fundamentally, all of our approaches to force-on-force analysis are underpinned by theories of combat that include both how combat works and what matters most in determining the outcomes of engagements, battles, campaigns, and wars. The various analytical methods we use can shed light on the performance of the force alternatives *only* to the extent our theories of combat are valid. If our theories are flawed, our analytical results are likely to be equally wrong. This is why some critics are skeptical of dire global warming predictions. The predictions are based, they say, on climatological models built on a very imperfect theory of how the Earth's climate works. For this reason, it is important to consider how our theories of combat are developed and where they come from.

Combat is an exceedingly complex simultaneous interaction of many factors. Large institutions, here and abroad, have been engaged for many years trying to develop theories of combat powerful and reliable enough to permit accurate predictions of combat outcomes. The most successful theories have been those for predicting the effectiveness of individual systems in combat in which most of the data comes from the physical realm (velocity, penetration, rate of fire, mean time between failures, etc.). We are reasonably confident that we can predict how a particular missile or radar will behave under different operating conditions. Our confidence falls rapidly as we try to forecast the results of more complicated combat situations that depend on the interactions of many weapons and units over time and that are critically shaped by human behavior and decision making. This does not mean that force-on-force analysis is useless for problems other than hardware selection. It does mean that we use the results fully aware of their limits.

CASE STUDY: MAJOR LEAGUE BASEBALL

Suppose you were the manager of a major league baseball team. You would execute a form of force-on-force analysis nearly every day. As you prepare for next season, you analyze the strengths and weaknesses of your team in terms of offense and defense, the pitching staff, fielding, hitting, and base-running ability. You also take into account similar information about your opponents, weighting most heavily the strengths and weaknesses of the other teams within your division. Data abounds, and you may have a considerable chore deciding which statistics to use. But, in general, the data would be reliable and easily available. From your analysis, you develop your trading strategy, shape your minor league teams, and modify your coaching staff.

During the season, play-offs, and throughout the championship series, you assess the opposition in detail and make near-term decisions such as which pitchers start which games and how you will rotate them. During each game, you decide tactics: whether to relieve a tired pitcher, walk a batter, insert a pinch hitter, order a bunt or a sacrifice fly, etc.

At each level—strategic, operational, and tactical—you grapple with the same issues as an executive decision maker in DoD. It's all force-on-force analysis. Which of the many pieces of data are meaningful (valid) and allow you to forecast the future? At what point does your method of force-on-force analysis become so complex and burdensome that its reliability and practicality suffer? You could develop an analytical approach so elaborate that it required the entire season and the team's payroll to run just once. The result might be quite valid but of no use at all.

Force-on-force comparisons cannot accurately predict who is likely to win or lose an engagement, a battle, a campaign, or a war. But we can use them with more confidence (though far from certainty) to predict whether one system, tactic, force structure, or course of action is likely to perform *roughly* better or worse than another. We can also use force-on-force analysis to assess why one alternative performs better than another and what happens to that performance when we change the forces, how they are used, and the conditions in which the combat occurs. But we must always keep in mind that if a force-on-force analysis embodies a completely erroneous theory of combat, even these more modest predictions are likely to be completely erroneous as well. That is why most of the complaints about the adequacy of force-on-force analytical methods, especially those involving mathematical models, are actually about the weaknesses of our theories of combat, especially future combat. This is the argument of many proponents of the Revolution in Military Affairs—that force-on-force analysis, as done today, improperly represents new weapons, technology, and operational concepts, thereby slowing their introduction into operational units.

Methods of Force-on-Force Analysis

Next, we will survey the most common force structure analysis methodologies used in defense resource allocation. They vary in their complexity from very simple order of battle comparisons to highly interactive dynamic models requiring tremendous computing support. Our previous caveat remains in force: computers have made it easier to model much more complicated theories of combat, but they can do no better than the theories of combat that drive them.

STATIC, SYMMETRIC COMPARISONS

Static methods are so-named because they exclude time. They are snapshots of aspects of the combatants we think are predictive of combat results. Usually, we express static measures as numbers and the difference between them or their ratio is taken to represent the superiority of one side over the other. The most straightforward use of static measures is the symmetric comparison, often referred to as a "bean count." Suppose we are Blue force planners tasked with assessing whether Blue forces are sufficient to defend against attack by Orange forces. This is a classic problem for which force-on-force analysis is used. A static symmetric comparison counts Blue tanks against Orange tanks, Blue aircraft against those of Orange, Blue troops against Orange troops, and so forth. To interpret what these comparisons mean for combat, we convert them to ratios. For example, a military rule-of-thumb since the days of Napoleon says that, to carry out a successful ground attack, an attacker must have a 3 to 1 advantage over a defender. For our question of whether Blue is at risk from Orange attack, as long as Blue prevents Orange from attaining a 3:1 advantage, we can defend ourselves against Orange attack. Note the theory of combat embodied in static, symmetric measures: in combat, like forces fight like forces, and the force ratio predicts the outcome.

What are the strengths and weaknesses of this approach? Its greatest strengths are its ease of use and transparency. Anyone is able to clearly see what is being compared and how. The data behind the static measures are usually readily available, and the mathematics involved is usually simple, so the measures are very reliable. Also, we know there is at least an element of truth in the theory of combat they embody. Numbers do matter. But how much do they matter and are they all that matters?

The weaknesses of these measures flow from the weaknesses in the theory that underlies them: quantity is not the only thing that matters in combat. In fact, there are many situations in which quantity may be the least important factor in determining the result of combat. Quality matters a great deal—great numbers of combat ineffective troops are irrelevant. Since World War II, the U.S. has chosen strategies that emphasize precise firepower (quality) over quantity. Logistics matters. So does command and control. Morale or generalship may dominate everything else. Also, modern combat is a combined-arms activity. Tanks fight infantry and anti-tank weapons in addition to other tanks. Artillery engages tanks, infantry, and other artillery as well as anything else in range. To remedy these weaknesses, we must add complexity.

Qualitative Differences

The first improvement we can make in the static, symmetric bean count is to account for the obvious qualitative differences between symmetrically arrayed forces. For example, we can modify the purely quantitative comparison by a multiplier, which represents the relative quality of the forces being compared. Usually we select one weapon or unit as the base (valued at 1.0) and compare the others to it. For example, comparing U.S. and Russian tanks and fighters, we may decide:

1 **M1A1** tank = 1.4 **T-80** tanks and

1 **F-15** fighter = 3 **MiG-25** fighters

These ratios reflect the qualitative edge that superior weapons give a force in combat; in this case we are saying that 1 M1A1 is worth 1.4 T-80s and 1 F-15 is worth 3 MiG-25s. We base these estimates of the relative quality on field data, professional judgment, and empirical evidence from laboratory comparisons. What began as a static, symmetric comparison of Blue versus Orange forces, based on quantity only, can be both quantitative and qualitative. So, if Orange has 300 M1A1 tanks, Blue can buy 101 M1A1 tanks or 141 T-80s (whichever option is less expensive) to prevent Orange from developing a 3:1 advantage.

What are the strengths of this approach? Obviously it represents a more complete theory of combat. Now that we have included quality, the validity of our static force-on-force analysis has improved; we more accurately model the real world. Unfortunately, qualitative comparisons are often subjective and difficult to make. Experts often disagree over the importance of a particular aspect of a system's performance and its proper weight when they establish a quality rating. For example, how important is the top speed of a fighter aircraft? Experts disagree and much depends on how you envision the aircraft will be used, itself an uncertain judgment.

This means that the numerical multipliers representing quality are difficult to agree upon, raising issues of reliability (are we measuring accurately?). How certain can we be that an M1A1 tank is actually 1.4 times better than a T-80 vice 1.3 times better, or twice as good? If F-15s killed MiG-25s at a three to one ratio during Red Flag exercises in Nevada does that mean they will achieve the same results in real combat… or was this kill ratio due in large part to some other factor such as crew proficiency? Our aggressor squadrons are generally far more proficient than our likely adversaries are. We know combat conditions also affect the importance of quality. There is some terrain where an M1A1 is worth at least ten T-80s, e.g., on the defensive with prepared firing positions, at night, and with long, open fields of fire. As we add more quality factors to better reflect the complexity of modern combat and improve validity, reliability declines as we introduce more measurement error.

Intangible Factors

Even if we have properly evaluated the quality and quantity of the weapons on each side, we still have not included some of the major factors that determine combat results. Many military commanders believe training, morale, unit cohesion, leadership, and generalship do more than anything else does to determine the combat effectiveness of a force. How can we incorporate these into static measures? Usually, we can use the same process we use for quality—we can apply a multiplier. The multipliers we use for morale, command and control, logistics, impact of casualties, etc., depend heavily on our theory of combat. An expert panel using whatever data is available chooses a number to capture the intangible capabilities of the two sides. For example, in the 1991 Gulf War with Iraq, many analysts compared U.S. and Iraqi ground forces. Because of many qualitative and intangible differences, the numerically smaller U.S. and coalition forces developed a far greater than 3:1 effective combat power superiority against their Iraqi foes.

The strength of this approach is that it improves the validity of modeling combat by incorporating more of the factors that we believe determine combat outcomes. The weakness is that numerical estimates of intangibles seem even more difficult, dubious, and unreliable than those for quality.

STATIC, ASYMMETRIC COMPARISONS

As we noted earlier, we need to ask whether it is valid to compare tanks against tanks, aircraft against aircraft, and so on, only symmetrically. Or, should we compare them asymmetrically by counting tanks against anti-tank systems and aircraft against air defense systems? Undoubtedly, introducing some sort of asymmetric comparison is appropriate because actual forces do not usually fight in a symmetric manner. In fact, one of our principal tactical objectives is to mass firepower and create a situation where we fight asymmetrically, strength against weakness, by bringing overwhelming combat power to bear at the point of attack, e.g., pitting all of our anti-armor systems— armored fighting vehicles, attack helicopters, close air support and supporting fires—against the enemy armor force.

Figure 8-1 displays an asymmetric theory of combat for World War II weapons systems. The direction of the arrows indicates success, thus armor can attack and defeat infantry, artillery, and air defenses but can be defeated by anti-tank systems and aircraft. Introducing asymmetry surely adds validity to our force-on-force analysis. Unfortunately, just as before, the addition of another factor also introduces greater unreliability. One reason reliability declines is because we have to decide what to compare with what, and how. For example, tanks are simultaneously armor, anti-armor, antipersonnel, fire support, and even anti-aircraft weapons. If we seek to compare Blue armor versus Orange anti-armor, we would surely include the tanks of both sides. But when we compare Blue infantry versus Orange antipersonnel weapons, should we count Orange tanks again? And when we compare Blue helicopters with Orange anti-aircraft weapons, should we count Orange tanks (with air defense machine guns) yet again?

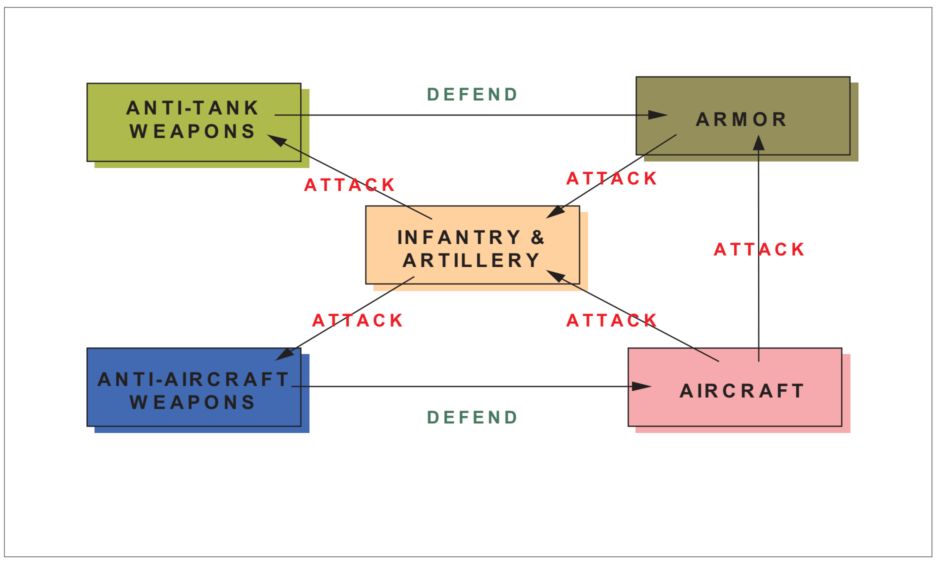


Figure 8-1. World War II Asymmetric Theory of Combat Model[1]

All multi-purpose weapons pose this problem of potential multiple counting. We must decide on the basis of the type of combat we anticipate whether a tank is best in one of its roles or another, or how to apportion it among them. A similar problem arises because only part of each side's force structure engages part of the other's at any given time. We distort this reality when, for example, we compare the total number of Blue tanks to the total number of Orange anti-armor systems. Different people would reach different decisions on these issues; hence this greater complexity introduces more reliability problems. It is difficult to reach consensus among experts about identical asymmetric force-on-force comparisons; as we cannot measure consistently.

SUMMARY OF STATIC COMPARISONS

We have seen how static measures can provide simple, clear snapshots of military capability at the price of limited validity: too much of what matters in combat is excluded. The remedy is to

1. Adapted from Archer Jones, *The Art of War in the Western World,* (New York: Oxford University Press, 1987), p. 611.

CASE STUDY: THE U.S. ARMY AND WEI/WUV'S

In the mid 1970's, the U.S. Army began a project to measure the difference in combat power among different types of Army divisions. Its earlier attempts focused on counting manpower in terms of Manpower Division Equivalents or MDEs. One MDE was a collection of 18,600 personnel in uniform. This calculation, easy to perform and hence reliable, was low in validity because it did not distinguish among types of equipment or tactical mobility, which are obvious determinants of combat power. Many leaders and analysts saw little utility in using MDEs as the basis for static comparisons. Next, the Army created a new unit of comparison called the Armored Division Equivalent, incorporating both manpower and quantity of weapon systems, although quality was still excluded. For this reason, the Army next developed a method for incorporating quantity and quality into static comparisons of ground forces.

The first step was to calculate, through laboratory and field testing and by convening panels of military professionals, a firepower score for each weapon in the Army inventory. These scores represented the accuracy and killing potential for every weapon against a standard target at a fixed range. Similar scores were developed to represent every weapon's mobility and vulnerability. By combining these scores, the analysts extrapolated a general capability index for each Army weapon. The result was a set of Weighted Equipment Indices, or WEIs, for each rifle, tank, cannon, etc.

Given WEIs for each weapon in a division, and the table of organization and equipment for each division, the analysts summed the WEIs to derive a single numerical score, called a Weighted Unit Value (WUV), for each U.S. division, our NATO allies, and our Warsaw Pact adversaries. For convenience, the score of the 2nd Armored Division, the heaviest and most powerful armored division in the world at the time, was given a score of 1.0 and other U.S. and foreign unit scores were set relative to it, e.g., the most capable Soviet Armored Division was given a score of 0.8.

The WEI/WUV methodology was very useful for comparing the combat effectiveness of units. While it was still a static measure, it was created using estimates of how units would perform in combat in terms of firepower, mobility, and vulnerability. Thus, it addressed some of the most important validity problems that afflict most static measures. The WEI/WUV method became a standard in DoD analytical communities and analysts used it as a starting point for wargames and umpiring battle results.

Unfortunately, the WEI/WUV approach had shortcomings. For instance, there was no WEI/WUV equivalent for close air support, yet air forces have a pronounced effect on the ground battle. Also, simply summing the capabilities of a unit's weapons did not capture the synergistic effects of the weapons operating together. That synergy, called combined arms by the Army, is central to operational concepts for conventional forces. Finally, intangibles such as training, morale, and leadership were not included in the WEI/WUVs. As a result, the Army has abandoned the WEI/WUV method in recent years as a force planning tool, although similar indices are still used as components in larger analyses.

include more of those missing factors at the price of decreasing reliability and practicality. As we add more complexity to static comparisons, we have seen how it becomes increasingly unclear exactly what the numbers mean that we seek to compare. When we began by comparing only quantity, the meaning of the numbers was completely clear. They represented the size of two forces. But each time we modify our numbers by including quality, intangible factors like morale and training, and then the asymmetry of combined arms, the harder it is to explain to ourselves and others what those once simple numerical comparisons actually represent and why our models work the way they do.

As we gain in validity, we lose the reliability that is the principal strength of static measures; we also lose transparency and reduce the ease of understanding the model, particularly when we are dealing with the uninitiated. The consequences of these weaknesses are severe. Either we give up trying to represent much of what we think matters in warfare (loss of validity), or we have to make the static measures increasingly complex, opaque, and arbitrary (loss of reliability and often practicality). Note that many of the factors that are hardest to incorporate in static comparisons are the very ones most important to us as we assess issues surrounding the Revolution in Military Affairs; e.g., the value and effects of very fast, fully integrated command, control, communications, computers, intelligence, surveillance and reconnaissance; increasingly speedy tactical decision making; and network-centered warfare.

Despite their weaknesses, the relative simplicity of static measures makes them attractive, and we still use them in DoD in many different forms. As you develop, use, and evaluate them, keep in mind that all the various methods of force structure analysis have strengths and weaknesses. None of them produce answers that are completely correct. Sometimes, maybe often, the validity problems that simple static measures suffer are tolerable. It depends entirely on the problem we are solving or the decision we are making. The speed and clarity of well-done static models may more than compensate for their limitations, e.g., they are still very useful for comparing nuclear arsenals.

DYNAMIC FORCE-ON-FORCE MODELS

The most fundamental limitation of static measures is that to remain simple they must exclude time. They freeze the capabilities of a military force at a particular moment. Yet we know that time and tempo are central to military operations, especially as we move closer to *Joint Vision 2020* operational concepts such as precision engagement and dominant maneuver. All the factors that static measures freeze change continually as military operations proceed. As forces maneuver, their new locations often change their capabilities. Similarly, their capabilities also change as they are reinforced and re-supplied, suffer attrition, expend munitions, and move in and out of contact with higher command. Successful military commanders use time more effectively than their adversaries do. Thus, no matter how sophisticated static measures become, they always (by definition) exclude time, a basic factor that determines the outcome of combat. Dynamic measures attack this problem.

Making a comparison dynamic means we incorporate the dimension of time. By better approximating the actual conditions of war, which are certainly dynamic, we increase the validity of the force-on-force analysis. But, as before, this added validity comes at the expense of reliability and practicality—which the use of computers has not overcome.

There are three general approaches toward making dynamic comparisons. The first type, mathematical models and computer simulations of combat, includes time in its mathematical representations of the forces and the variables at play. The second group, exercises and experiments, incorporates time by using the real forces themselves. The last, wargames, models time by having participants play the game in turns while using maps and demarcations, symbols in place of actual forces, and rules governing their behavior. Remember, just as with static models, whatever form dynamic force-on-force analysis takes and no matter how many factors are included, the results can never be any better than our understanding of what determines the outcomes of battle in the real world, our theory of combat. If it's true that dynamic model results cannot be better than the theory of combat embodied in that model, regardless of its computational power, what can we learn from such models? Sometimes we can learn a lot. To see this point, let's consider the case of computer-driven dynamic models.

First, think about what people do well, and what computers do well. Computers are good at keeping track of large volumes of input data, and at consistently following the rules programmed into them. They will tell you exactly what is implied by input data, given all the rules you have programmed in about how to manipulate that data. And they will do so even when any person would recognize right away the results don't "make sense." In contrast, people are good at seeing the big picture, placing facts in context, and being imaginative. On the other hand, human beings are not always consistent—we might say that we believe X, Y, and Z, but nonetheless shy away from implications that don't "make sense" or are otherwise unacceptable.

Given these relative strengths and weaknesses, we can learn from dynamic computer models even though we are the source of everything that those models "know." This is especially the case if the model produces results that aren't what we expected. In that case, the model is telling us that, if we believe input data values A through Z, and if we believe rules "A" through "Z" for manipulating those values do a good job of stating our theory of combat, then we ought logically to accept the model's conclusions. If we don't, it doesn't mean we are wrong and the model's right. (Far from it.) But it should force us to think. Is there something wrong with the input data or the manipulation rules that embody what we said was our theory of combat? Or, given that data and those rules, ought we to accept the model's results?

Other kinds of dynamic modeling can also force us to think. If exercises, experiments, and wargames are honestly run, with players given free rein to do what they see fit, then these dynamic simulations can also produce unexpected results. In trying to explain those results, we can gain insights that help us question the "conventional wisdom" and refine our theories of war. Below, we discuss each type of dynamic simulation in turn.

Mathematical Models and Computer Simulations

Mathematical force-on-force models have grown increasingly sophisticated as computing power has become more available. Generally, the more abstract arrangements tend to be referred to as mathematical models, while the more detailed and complex assemblages are called simulations because they are supposedly more life-like or closer to experiential reality. These models range from very rigorous engineering representations of individual items, through system (often vehicle) simulators that we often network to one another, to more aggregated but still highly complex models of theater military operations. Theater-level combat models almost always play an important role in major DoD resource allocation decisions because such decisions are ultimately aimed at improving combat effectiveness.

Although mathematical models are used to assess all levels and types of warfare, we will focus on the strengths and weaknesses of the theater-level models used by the Department of Defense. They are, at heart, elaborate pieces of software that contain mathematical representations of the aspects of theater operations their designers deemed important to determining theater campaign outcomes. These include air, ground, and sea forces; logistics; weapons of mass destruction; command, control, communications, and intelligence; morale; and strategic lift. In theater-level combat models, the areas of operation are represented as maps with icons representing forces and units. Operations are conducted in blocks of time, sometimes variable. These models usually can be run fully automatically but they also can be stopped at any point, reversed, modified, and rerun, making them ideal tools for exploring branches and sequels and for sensitivity analysis.

In any theater-level combat model, all the air, ground, and sea units and their individual characteristics have to be loaded into the computer, unit by unit, before we can use the model. These data are inserted into an enormous series of spreadsheets. Ground units are usually represented at the brigade level, air units by squadrons or wings, and sea units by battle group, although most models permit operators to use greater or lesser aggregations. Each side's concepts of operations or decision logic must be loaded into the model so that every unit reacts to each eventuality. Not surprisingly, these instructions require constant adjustments since the range of eventualities is so great. In sum, the preparation of the model for use is labor-intensive and full of opportunities for errors that are discovered only by trial and error—if at all.

The most widely used theater-level combat model in DoD is the Tactical Warfare Model (TACWAR), managed for the Joint Staff by the U.S. Army Training and Doctrine Command, although each of the services also has its own models. The Office of the Secretary of Defense and the Joint Staff use TACWAR extensively to examine force planning options. The unified commanders test their Operational Plans using TACWAR. We will use TACWAR, as DoD used it during the Deep Attack Weapons Mix Study (DAWMS) in 1998, as an example for this section of the text.

AIRCRAFT	MK-82	CBU-87	AGM-65	GBU-27	JDAM	JSOW
AV-8B	6	4	2		4	
B-2	80	24			16	16
F-16C	6	4	4		4	2
F/A-18C USMC	12	8	8		4	2
F-117				2		

Table 8-1. Allied Aircraft Weapon Payloads.

Table 8-1 is a partial representation of some of the combinations of aircraft and payloads analysts inserted into TACWAR preparing for the Deep Attack Weapons Mix Study. Note the assumptions the analysts have to make to create a spreadsheet like this: we know these aircraft are capable of carrying these loads, but realistically would they? Are these representative loads for a typical mission in this theater? We know that if the aircraft is based closer to the target, or does not need to loiter waiting for a call from ground forces to strike, it can carry more bombs and less fuel; what did the analysts assume for these aircraft? Also, they do not allow mixing weapons types, etc. We use the data in these tables to represent all the aircraft of that type in the simulation, so the difference between loading four versus six bombs on an AV-8 has important implications for the "worth" of that aircraft and how it contributes to building combat power.

AIRCRAFT	SURGE SORTIE RATE	SUSTAINED SORTIE RATE
AV-8B	3.5	2.53
B-2	0.8	0.55
F-16C	2.5	1.96
F/A-18C USMC	2.9	2.24
F-117	1	0.57

Table 8-2. Allied Aircraft Sortie Generation Rates.

Combined with table 8-1, table 8-2 gives us a clearer sense of how TACWAR works and its theory of combat. The payload and number of missions each aircraft flies (a sortie is one mission flown by one aircraft) together generate combat power; the most powerful aircraft carry more weapons and fly more often. The surge sortie rate is the maximum possible number of missions in a 12-hour TACWAR cycle the aircraft can fly during an emergency, e.g., during the Halt Phase when a breakthrough or overrun of friendly forces or key terrain seems imminent. The sustained sortie rate can be maintained more or less indefinitely.

AIRCRAFT	CLOSE AIR SUPPORT	STRATEGIC TARGETS	SAM SUPPRESSION	GROUND-CON TROLLED INTERCEPT SITES	AIR BASE ATTACKS
AV-8B	1.00				
B-2	.6	.2	.2		
F-16C	.3	.05	.4	.15	.1
F/A-18C USMC	.45	.2	.25	.1	
F-117		.4	.3	.3	

Table 8-3. Allied Aircraft Target Allocations.

TACWAR needs to know the missions and target sets the planners will allocate to each type of aircraft. Table 8-3 shows the distributions of effort for the aircraft from the earlier tables that were used in DAWMS. For example, 60 percent of B-2 missions will strike enemy ground forces invading the nation we are defending, 20 percent will attack strategic targets like power grids and command centers, and 20 percent will attack enemy Surface-to-Air Missile batteries. To build this table, the military planners must make operational choices about the overall air campaign for the theater and then set the level of effort for each different target set. The planners must also decide whether they will change their apportionment during different phases of the campaign. TACWAR has five Attack Mission categories, some with as many as four sub-categories, and two Defense Mission categories (Battlefield Defense and Area Defense).

Just as we discussed earlier with static combined-arms models, multi-mission aircraft pose a problem for the analysts. Their roles may actually be situational, dependent on enemy actions and levels of activity. Analysts can try to accommodate these actions in the model with a series of "If… then" rules, but they do so by introducing yet more complexity and they require extensive help from operators to ensure they use reasonable rules.

Apportioning aerial effort is where service cultures and doctrines clashed so mightily during the Gulf War: does the CINC or his/her J3 make this apportionment decision or does the Joint Forces Air Component Commander (JFACC)?[2] How do we account for excess sorties the Navy and Marines will provide after they have met their own requirements? How many sorties of

2. Current joint doctrine has the Joint Forces Air Component Commander propose the apportionment to the CINC. This allows the other Component Commanders a built-in opportunity for reclama if they disagree with the proposed apportionment.

what type will halt an aggressor without sacrificing key terrain? What if resources are limited and the CINC needs more sorties than the maritime services volunteer? While many of these answers are rooted in doctrine and service procedures, TACWAR can tell us what answers we ought to be prepared to accept, if we agree with the data values and other assumptions input into the model. For example, in an otherwise fixed scenario, what happens if a Navy air wing reduces its counter-air defense combat air patrols (interceptors) and dedicates them to ground support? Is the enemy's halt line significantly altered? Does the carrier or its escorts become unacceptably vulnerable?

Another important preliminary step the analysts must set into TACWAR is a map of the theater of operations. After the map boundaries are set based on the scenario we are examining and the physical terrain is input and verified, the analysts identify key military terrain and facilities that affect both sides, e.g., aerial and sea ports of debarkation, roads, bridges, urban centers, bases, economic objectives, etc., as shown in figure 8-2.[3]

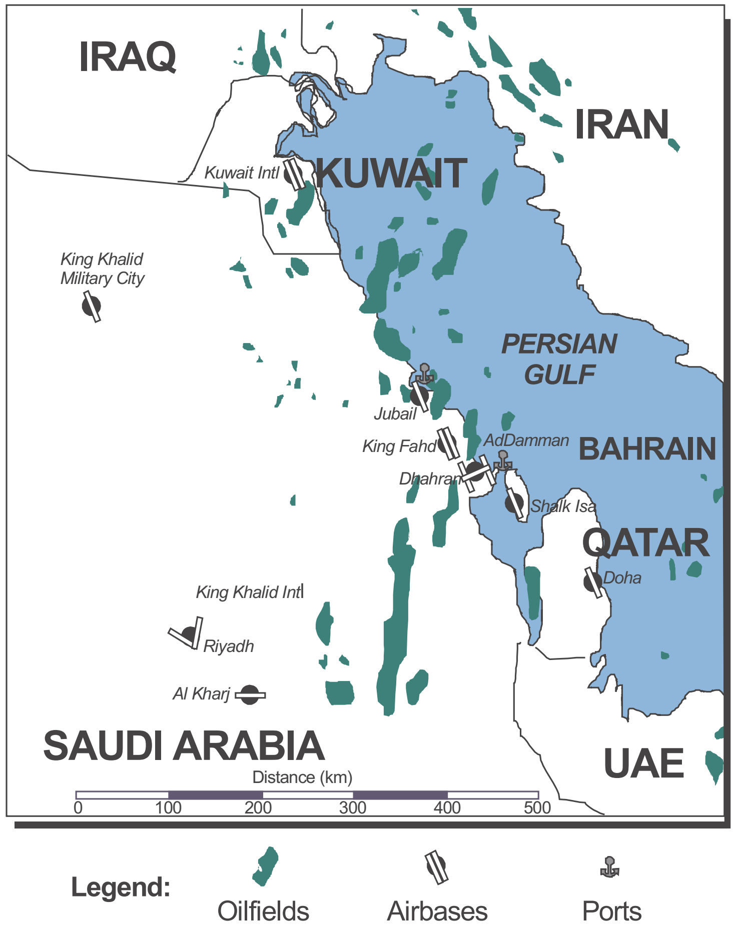


Figure 8-2. Notional Key Terrain in Southwest Asia.

Theater-level combat models conduct their campaigns by moving forces to their objectives; along the way they may make contact with the enemy. Generally, ground units advance along scripted axes. As shown in figure 8-3, the sectors (or cylinders) vary in shape and size depending upon how they conform to terrain. The analysts often place more numerous, smaller sectors in areas where they anticipate contact between opposing forces.

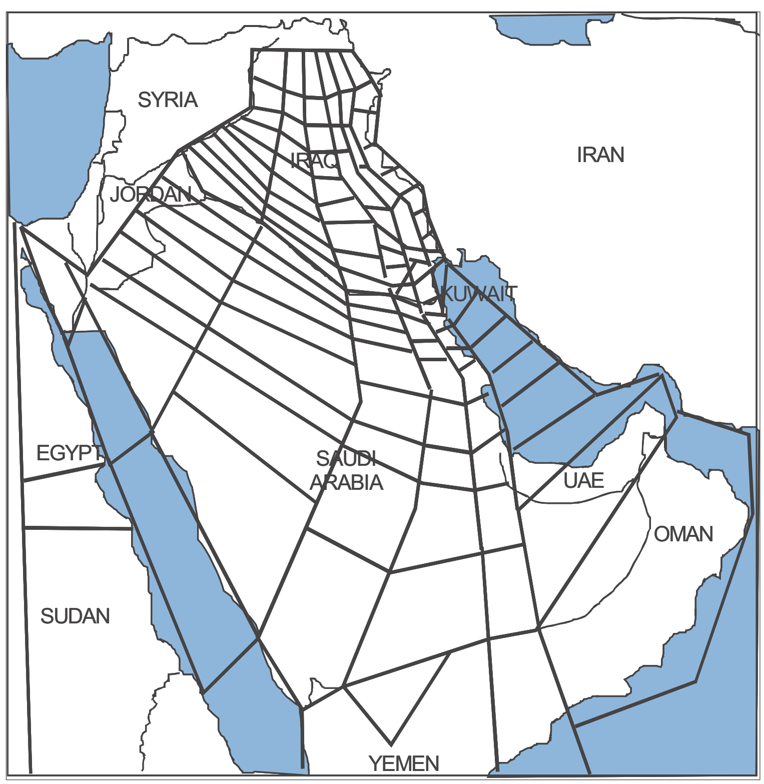


Figure 8-3. Notional Sectors Overlaid on a TACWAR Map.

3. The TACWAR figures presented here are labeled as notional because they represent a hypothetical (illustrative) scenario versus one prepared or confirmed by intelligence officers, logisticians, and other staff planners for actual operational planning.

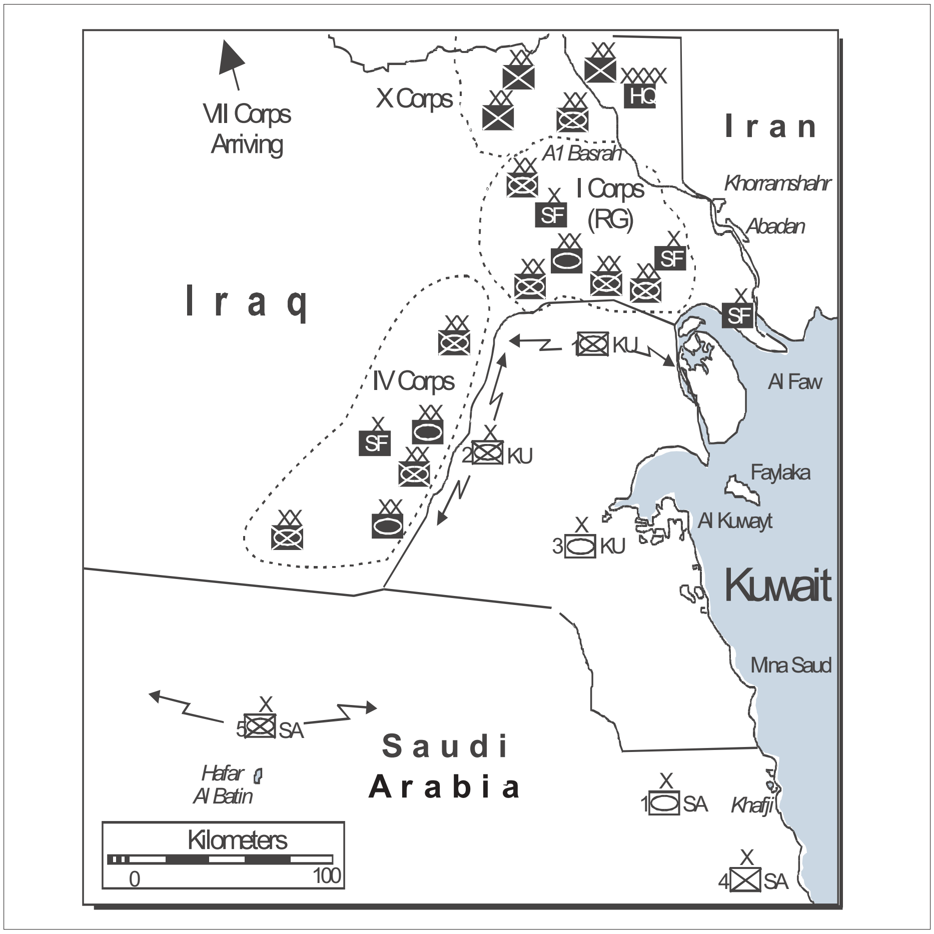


Figure 8-4. Notational Disposition of Allied Ground Forces

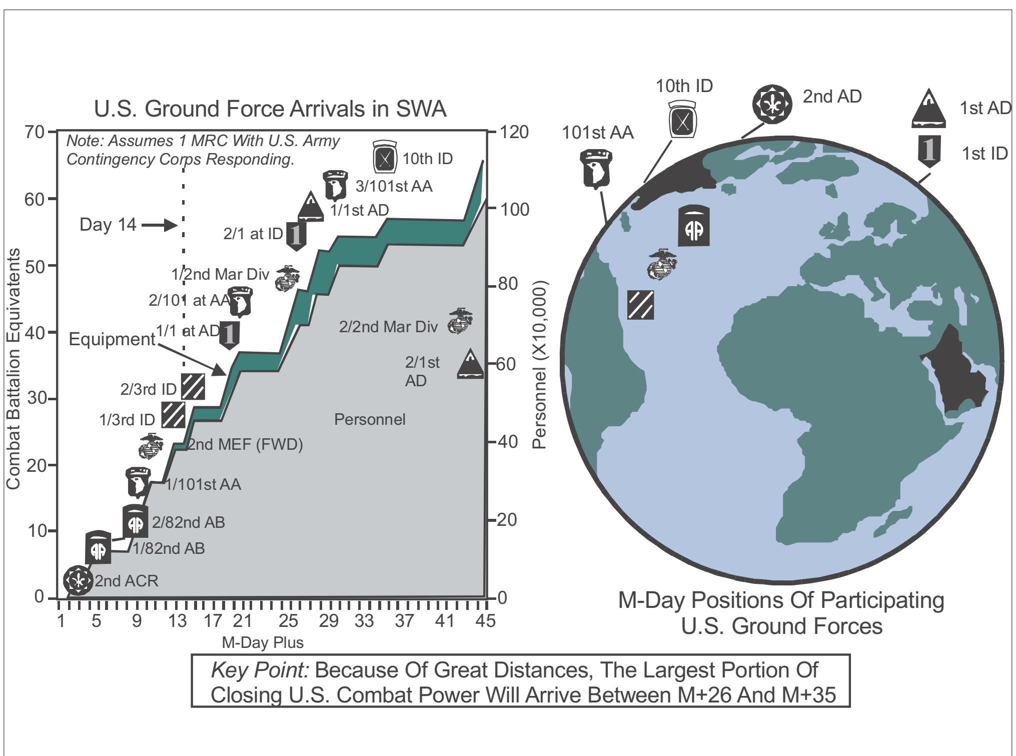


Figure 8-5. Notional Arrival of U.S. Ground Forces.

There are other imaginary lines in TACWAR, too, that the analysts can adjust, e.g., when does close air support of ground troops stop being close air support and become battlefield interdiction? How far behind the battle lines can deep strikes reach?

Now that the map is set and the analysts have specified the capabilities of individual units, they create the initial disposition and flow of forces into the theater. These are variables that may be central aspects of the analysis. For example, how much earlier is the enemy halted (according to the model) if we have three, rather than two, Army brigade sets of equipment pre-positioned in Southwest Asia? The Joint Staff Mobility Requirements Study 2005 used this kind of TACWAR modeling to find the smallest, latest arriving series of forces that halted the invaders short of key terrain with moderate risk. Figure 8-4 shows the initial disposition of Allied forces used in DAWMS for its Southwest Asia scenarios.

Figure 8-5 displays the flow of U.S. ground forces, listed by brigade and measured in battalion equivalents, into Southwest Asia in the Deep Attack Weapons Mix Study. TACWAR allows the strategic lift of ground units to be separated into personnel and equipment because they arrive by different means, airlift and sealift respectively. As used in DAWMS, TACWAR assumes units are ready to move and standing by for strategic lift, i.e., it ignores complicating issues like disengaging from current operations like Bosnia or Kosovo, reconstitution, training, fort-to-port movement, and (for DAWMS) en route attrition.

TACWAR analysts must go through a similar process for aircraft, including arrival times and especially bed down in theater. Figure 8-6 shows the build-up of U.S. airpower from mobilization day forward by aircraft type. Note the rapid availability of long-range bomber aircraft and the steep ramps upward as each aircraft carrier arrives.

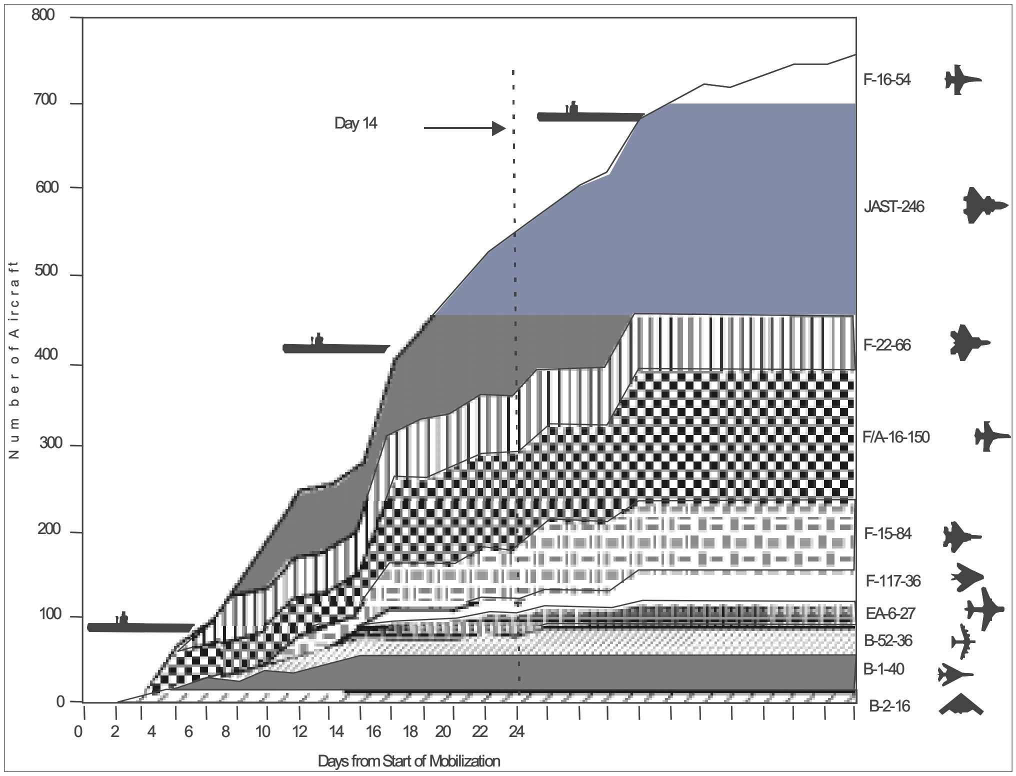


Figure 8-6. Notational Arrival of U.S. Air Forces.

All the land-based aircraft that flow into the theater must be bedded down at air bases with sufficient capacity, as we show in figure 8-7. Here is where coalition planning is especially important. While all national planners know the Maximum On Ground capacity of each base, they must coordinate to ensure that collectively they do not exceed it.

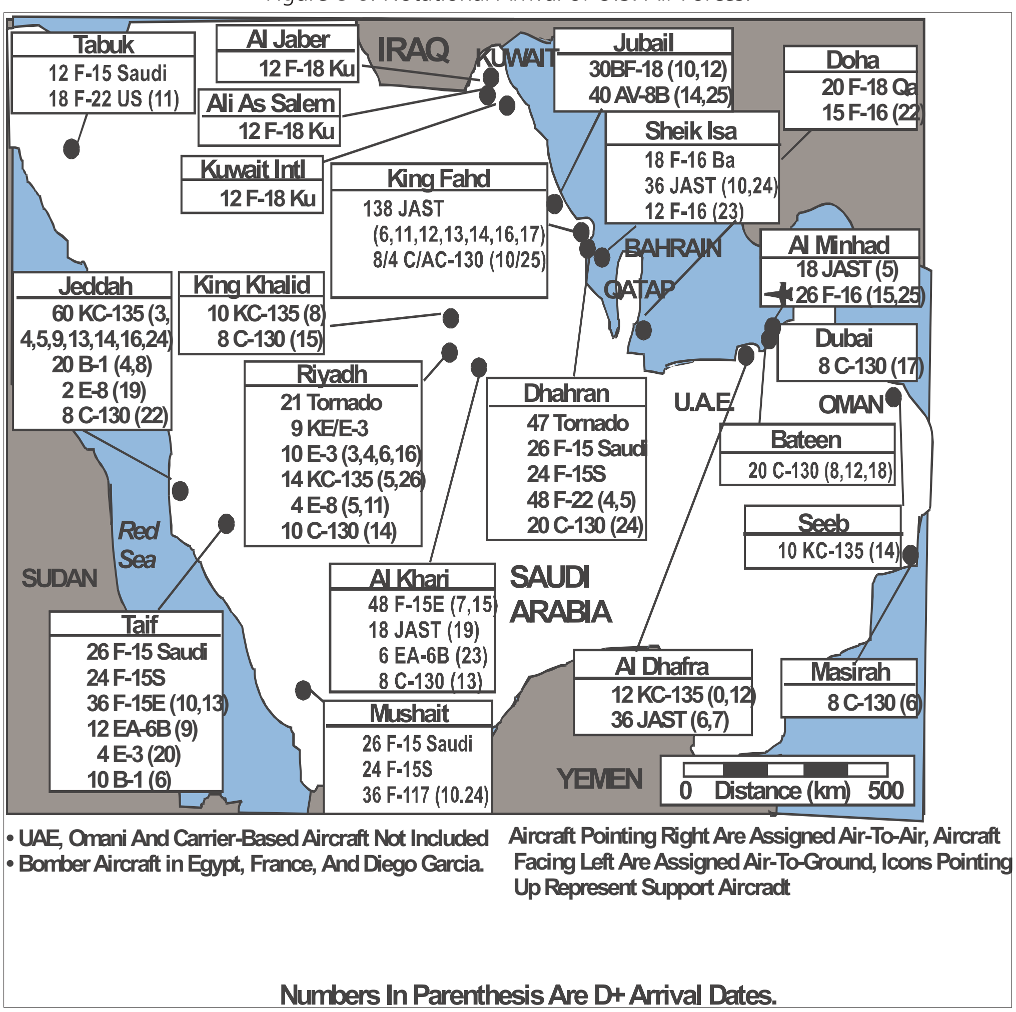


Figure 8-7. Notional Allied Air Basing in Southwest Asia.

Having identified the capabilities of individual weapons and units, and now their quantity as a function of time, i.e., initial dispositions and reinforcements,

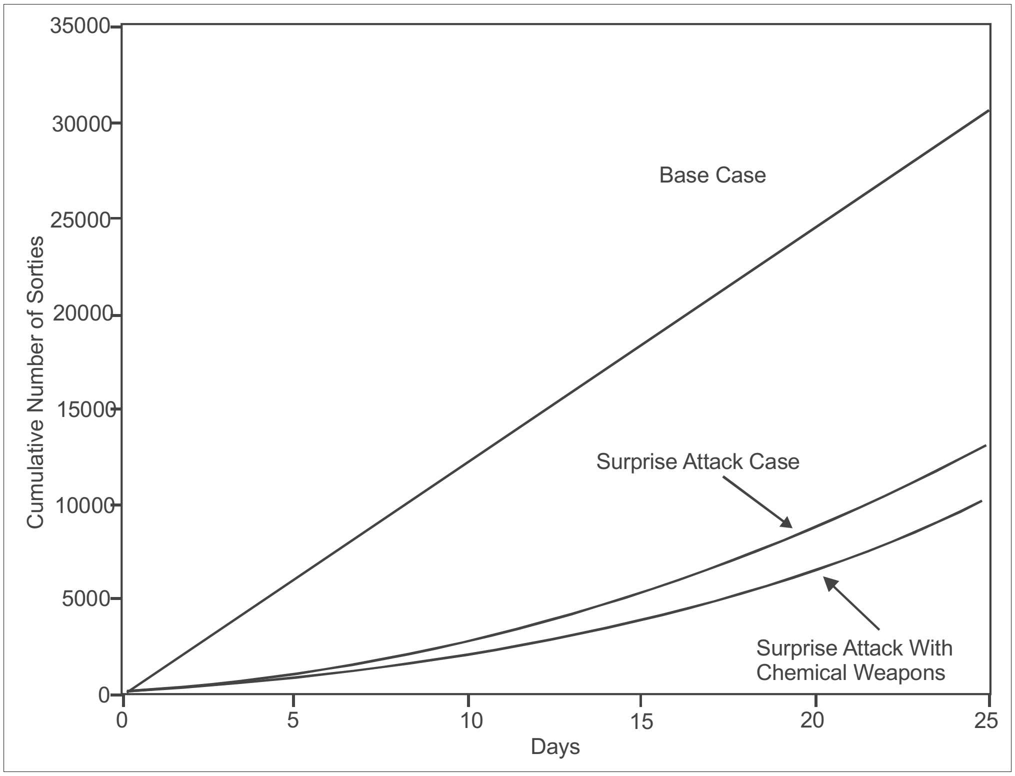


Figure 8-8. Notational Allied Sortie Generation

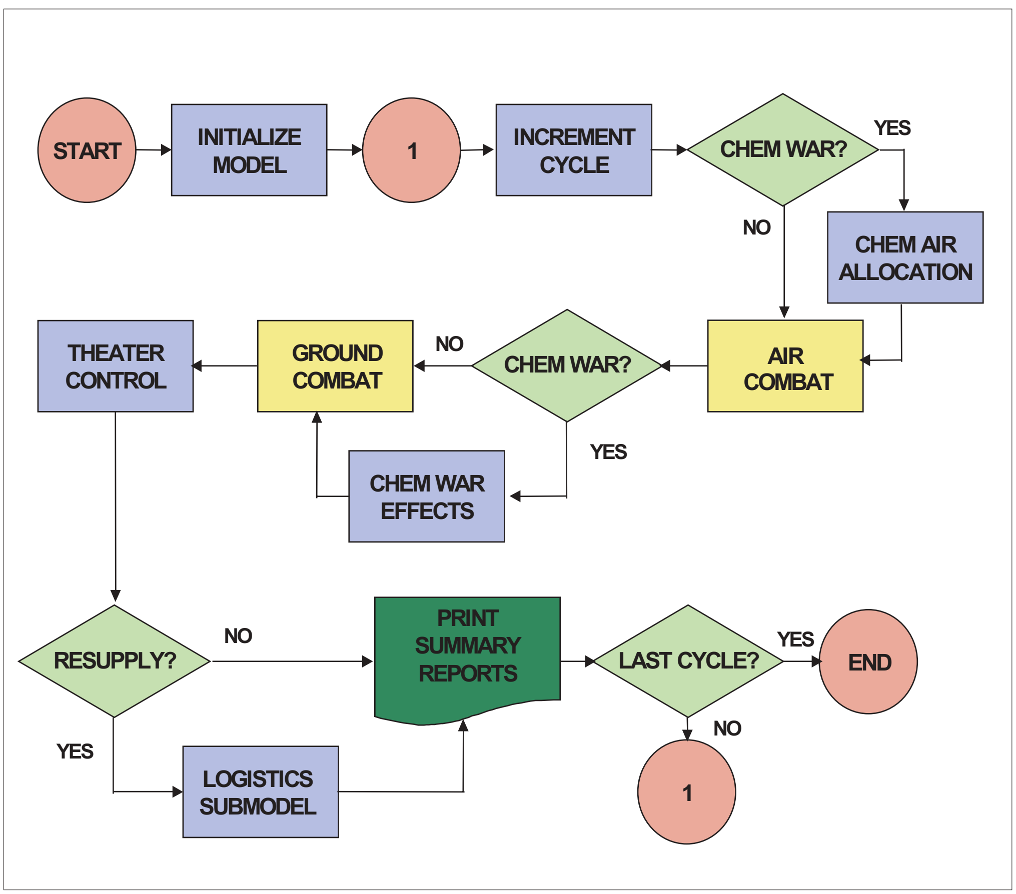


Figure 8-9. TACWAR Model Flowchart[4]

the analysts using TACWAR can calculate how quickly combat power builds up. They can generate charts like figure 8-8 to display aircraft sortie generation and display the sensitivity of this variable to changes in assumptions. The hypothetical base case in DAWMS assumed the Coalition Partners would have advanced warning of an Iraqi attack and that they would refrain from using chemical weapons; this graph shows the impact —according to TACWAR's theory of combat—of a surprise attack and the use of chemical weapons on sortie generation.

The modelers perform a similar analysis of the opposing forces—their initial dispositions, capabilities, likely axes of advance, reinforcement rates, etc., and then they are ready to run the model. TACWAR advances the ground forces along the cylinders until they have moved as far as they can in the cycle or until they make enemy contact.

Figure 8-9 shows the sequencing in TACWAR during each simulated 12-hour cycle. First, the model assesses the effects of the optional Chemical Warfare module and applies modifiers prior to air

4. Based on Figure 0-2 from Steve Kirin's "Executive Summary," *TACWAR Integrated Environment* (U.S. Army Training and Doctrine Command Analysis Center).

combat. Air operations are handled abstractly but within the model and prior to resolving ground combat, which may also be affected by chemical warfare. In TACWAR, players move ships toward and into operating areas or allocate them to patrol routes, but naval movement and combat is handled off line, outside the model. Naval forces inject firepower into sectors much like strategic air power, i.e., without the limitations and vulnerabilities of on-map basing.

TACWAR calculates air superiority within each cylinder. Based on the instructions the analysts provided during the set up, TACWAR assigns counterair units to cylinders in which their strength and capability is compared to the enemy counterair presence. Doctrinally, behind friendly lines, this is Defensive Counterair; in front of them it is Offensive Counterair. TACWAR calculates losses for each side and leaves the residual counterair capability of the superior force in the cylinder for the rest of the cycle. This residual counterair force may have the opportunity to engage strikes and ground support aircraft (and their escorts) in their cylinder based upon their remaining weapons.

Strike and bomber air units "fly" to their targets through the cylinders and, depending on their mission and profile, they may be subjected to attrition from surviving enemy counterair—interceptors and surface-to-air missiles. Support and escorting aircraft such as fighters and electronic jamming aircraft may negate some or all of enemy air defenses. TACWAR then calculates the surviving combat power's effect on their target sets and the users can request TACWAR results as we show in figure 8-10. The graphs indicate the Coalition's reduction of Iraqi Ground-Controlled Intercept and Surface-to-Air Missile battery air defenses over time in the base scenario and two sensitivity variations.

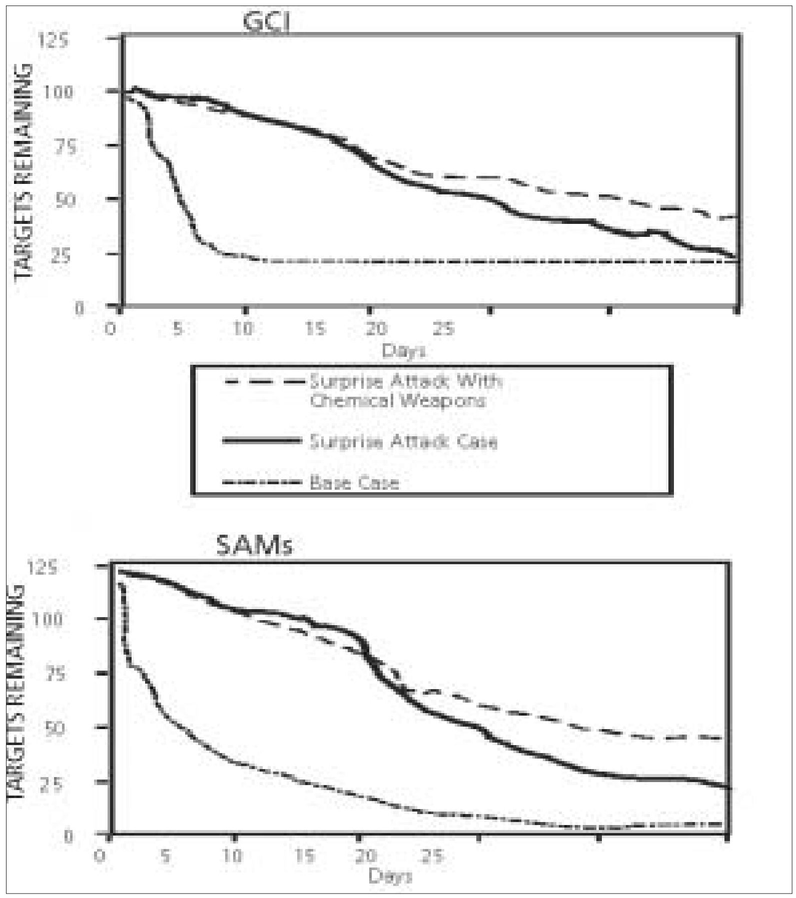


Figure 8-10. TACWAR Target Attrition.

The ground forces and their interactions are the original design focal point of TACWAR; many of its features such as expanded air warfare, logistics, and chemical weapons were added later to improve its validity at modeling modern warfare. TACWAR moves units, has them make various kinds of attacks, or conduct various kinds of defenses based on the instructions of the analyst and the participants. When opposing forces occupy the same cylinder, TACWAR calculates whether either side has enough force superiority to attack (regardless of the overall tactical situation) and, if that side is ordered to assault, it compares the combat power of the forces, including close air support, *in each sector individually*. The threshold required to attack is set by the analyst and may be adjusted between cycles and varied for each side.

The tactical posture of the units affects combat results, e.g., units ordered to delay will fall back to minimize casualties, moving the Forward Edge of the Battle Area (FEBA) with

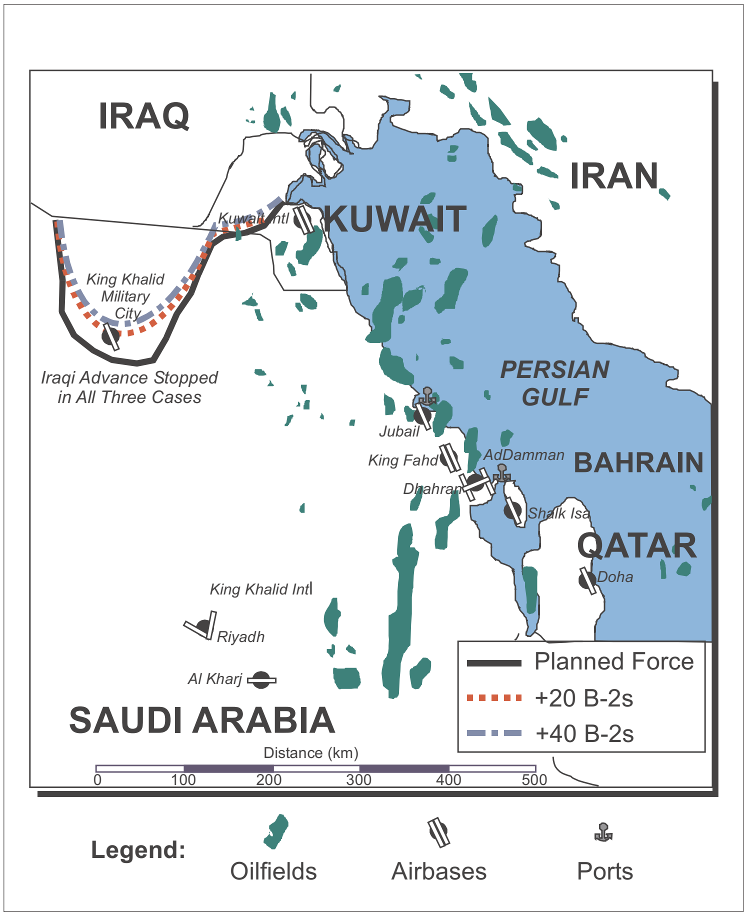


Figure 8-11. Battle Lines Calculated by TACWAR for DAWMS

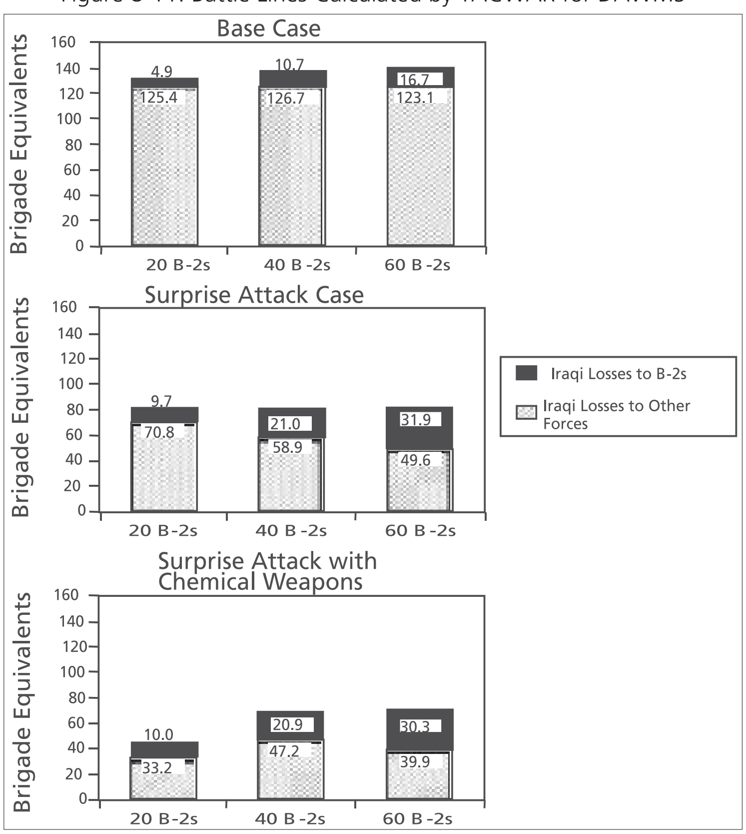


Figure 8-12. Notional Iraqi Ground Force Losses.

them as they fall back, whereas a unit ordered to hold will take greater casualties to prevent the capture of an objective and prevent movement of the FEBA. Depending on the combat results, the battle line between the opposing forces moves back and forth along the cylinder axis like a piston. Hence, models like TACWAR are often referred to as “piston-driven” models. TACWAR then links together and smoothes the piston positions of each cylinder to display the theater battle line after each cycle as shown in figure 8-11.

Note that figure 8-11 contains some sensitivity analysis. The Deep Attack Weapons Mix Study had a two-fold purpose: first, to identify the optimal mix and quantity of deep attack weapons among services and, second, to determine whether the U.S. should purchase additional B-2 bombers so as to place the order for more aircraft before the production line shut down. The different battle lines reflect the contribution of one and two additional increments of 20 B-2 bombers.

In order to establish how the battle line shifts, the combat resolution table determines the results of each enemy contact at the end of each time block in terms of casualties, logistics consumed, ground gained or lost, and targets destroyed. Those figures can be extracted after any cycle in the scenario, or at its conclusion as shown in figure 8-12 that shows Iraqi losses. These graphs are taken from the second part of DAWMS and reflect sensitivity analysis in both the quantity of B-2s and in the nature of the scenario. They demonstrate how tactical surprise and chemical weapons reduce allied effectiveness (fewer Iraqi losses) and how B-2s are relatively unaffected by either.

Figure 8-13 displays the reverse of the coin in figure 8-12, U.S. and Coalition Partner ground losses. They further demonstrate, as B-2 advocates would argue, that B-2s are increasingly valuable as the situation becomes more dire for the allies—nations who are becoming increasingly sensitive to casualties.

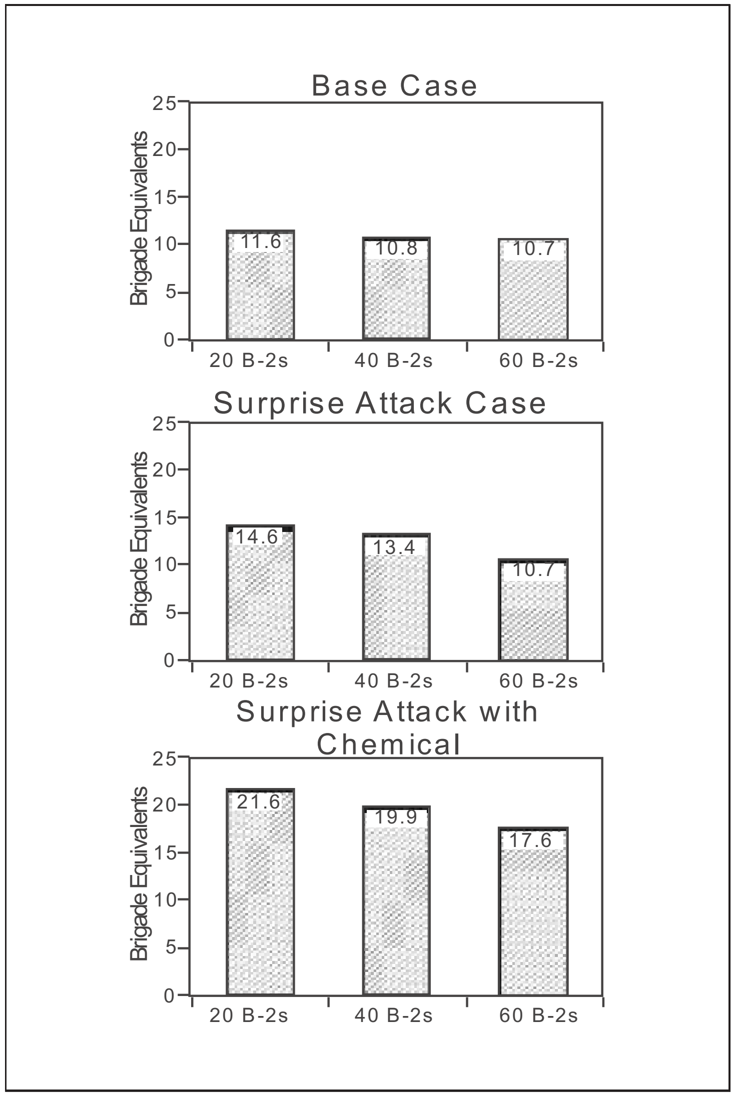


Figure 8-13. Notional Coalition Ground Force Losses.

The combat resolution table is one of the key components of any theater-level combat model. For TACWAR and the other theater-level models we use today, the combat result engines are collections of mathematical representations (equations and matrices), modified forms of Lanchester's equations adjusted with combat data gathered mostly from World War II and the Korean War. TACWAR begins combat resolution by aggregating the strength and quality of the opposing sides' weapons (much like the WEI/WUVs we discussed in our earlier section on static force-on-force methods) to calculate attrition in each sector.[5] It adjusts each unit's weapons' effectiveness based upon adjustments for logistics, training, chemical weapons, tactical posture, etc., and determines how many enemy personnel became casualties, how many enemy weapons were destroyed, how much ammunition and fuel was consumed during the 12-hour exchange of fire, and, as a result, how the FEBA shifted.

TACWAR also has several other modules that can be turned on or off or used to change data during an analysis. The Logistics Submodel overlays a network of supply points and places a hierarchical distribution grid over the sectors. Both are vulnerable to air attack. It monitors consumption and resupplies units in the priority order set by the analyst. The Theater Control routines control the interface with the map and its sectors that affect the level of detail or granularity in the scenario, i.e., smaller sectors require modeling smaller units. It monitors the battle lines and adjusts the boundaries of the rear areas like communication zones as the FEBA moves. The Theater Control Submodel assesses unit requirements, assigns replacements (weapons and personnel), and withdraws ineffective units. It calculates which airbases must be abandoned (if any) due to FEBA movement and advances units from rear areas toward combat sectors as directed by the instructions in the model or by the operators and analysts.

5. TACWAR uses the deterministic Antipotential-Potential method, a complex approach for assessing the value of each weapon toward destroying any other weapon and personnel it may engage with effect. Using a standard weapon as a reference or benchmark, analysts rate other weapons against it, building weapon-to-weapon kill matrices to determine ground weapon attrition. The engine aggregates the weapons in each sector to calculate how many opposing weapons of each type they destroy. Personnel effectiveness at operating the weapons is based on unit strength and logistics, modified by chemical warfare protective gear, training, etc. (Steve Kirin, "Executive Summary," TACWAR *Integrated Environment* (U.S. Army Training and Doctrine Command Analysis Center)).

Similar to the logistics submodel, TACWAR has a Command and Control (C2) Submodel. It overlays the sector map with a C2 grid and penalizes units' effectiveness by making fewer weapons available for ground units or generating fewer sorties if C2 is degraded. C2 effectiveness is degraded by unit casualties and by casualties to headquarters units in its chain of command. The limited Naval Submodel allows amphibious surface assault, i.e., fights to seize the beach. It treats aircraft carriers as floating airbases that generate sorties in support of the ground and air wars, and TACWAR models surface fire support like off-map artillery.

What are the strengths and weaknesses of theater-level models for force-on-force analysis? Their primary strength is that they are the only way we have that captures most of the aspects of combat we believe are so important. In this sense, they promise the greatest validity of all the approaches we have discussed so far. This validity advantage comes at the price of enormous complexity, and this is their weakness. This complexity is so great that it is very unclear (perhaps unknowable) how valid these models are at representing our theory of combat.

The reliability problem begins with the enormous amount of data and the many assumptions these models need to generate their output. Seemingly small changes in the inputs at any stage can produce disproportionate and unintended changes in the outputs. With so many inputs, we may not be able to isolate what is causing a particular result, especially if some obscure but sensitive detail of an assumption or piece of data is far upstream from the spurious result. This means that the results of an analysis can be highly sensitive to the decisions that the analysts make as they prepare the models—it can also mean that these models are subject to subtle manipulation in the hands of those who know what they are doing. For example, how will the analyst score the capability of a particular weapon? How many sorties will he permit a particular platform? How fast can tracked and wheeled vehicles transit a particular piece of terrain? What is the effect of weather on a particular suite of avionics? What is the margin of superiority required to attack? How many casualties can a unit sustain before it ceases to be effective in combat? How does an organization react when it is disconnected from its higher headquarters? How fast does combat power decline when a particular logistics node is interdicted?

The model per se does not tell us these things. Instead, it provides a platform for representing whatever values for these questions the analysts deem appropriate. It should be clear that, quite often, we have no objective way of knowing what is an appropriate value for answering such subjective questions. Thus, we can see how the services can use the same model in similar scenarios and generate different results and why they do not accept each other's analyses. We can also see why using these complex models is as much art as science.

Some have argued that an easy way around these problems exists. We could tune these theater-level models to imitate the results obtained in some real battles from World War II. The U.S. Army did this with its Concepts Evaluation Model, an Army-modified version of TACWAR that models ground war only. They ran it for the 1943 Battle of Kursk in Russia and the 1944 Battle of the Bulge in Western Europe. As one would expect, the results tracked some historical outcomes well and missed others badly. In particular, the model did a poor job of capturing the intangible factors that are so important in combat, especially morale. For example, historically, the behavior of two nearly identical Soviet units under similar combat conditions would vary widely and inexplicably; indeed, the same unit would vary its behavior from day to day. The modelers could not replicate or predict a pattern; the closest they could come was inserting random events, which was clearly unsatisfactory. In the Battle of the Bulge, German tank losses in the model were far higher than in the actual battle because it did a poor job of replicating the shock effect of an unexpected German attack on inexperienced U.S. troops.

But even if models could be made to repeat history perfectly, that approach is still not very useful. The pace of change in war is simply so great that we have little confidence that experience from previous wars is sufficiently relevant to justify setting modern theater-level models to replicate historical results. How different is combat today from combat in World War II? Also, many battle outcomes were the aggregate result of numerous low probability events and decisions. In June 1942, the Battle of Midway between the U.S. and Japanese navies was decided largely by series of tactical mistakes by commanders and intuitive decisions by small unit leaders in a sequence that is very unlikely to be repeated under any circumstance. When we tune a model to reproduce a historical result, we are in effect saying we believe those low probability events and decisions should be expected and incorporated into all our analyses and future conflicts. Plainly that is not appropriate.

Some argue that a piston-driven model based on Lanchester's Equations, such as TACWAR, is inherently incapable of representing modern warfare at the theater level. An attrition-based model cannot adequately reward maneuver let alone "effects-based warfare" that strives to paralyze an enemy's command and control of his forces and induce psychological and information warfare effects as well as physical damage. In short, an attrition-based model equates to a flawed theory of modern combat, say these critics. The counter-argument is equally simple: at some point any effect becomes discernible and affects the outcome of combat; therefore it can be included in the model—if we can agree upon the nature and the magnitude of the effect.

Like any other analytical method, a mathematical model can do no better than the theory of combat that it is intended to portray. The model cannot tell us how ground or air forces fight. It can only tell us that, given a particular theory of how they fight, a particular alternative is likely to produce a particular result. For this reason, modeling results can never be construed as a point prediction of what we can actually expect in the real world. However, we do use these models for weaker kinds of predictions, e.g., whether one alternative is likely to perform better than another in the real world.

Thus we must insist that all analyses using these force-on-force models assess the sensitivity of the results to changes in key variables and that they compare alternatives without making changes to them. The caretakers of our models should be comparing their outputs constantly with new information gleaned from actual conflict and from experiments. We should take advantage of the immense practicality of these models—once they are built, analysts can run them many, many times at little additional cost—to root out inconsistent outputs from small changes. Once again, we must never treat the output of a model, no matter how sophisticated, as something to be taken at face value. The model is never responsible for its own results; the users of the model are and they must analyze those results keeping the limitations of the model in mind.

CASE STUDY: PREDICTING THE OUTCOME OF DESERT STORM

Immediately prior to the beginning of the air campaign in the 1991 war with Iraq, Dr. Joshua Epstein of the Brookings Institution released his force-on-force analysis of the impending conflict based on his Adaptive Dynamic Model. He forecast some 16,000 American and Allied troops would be wounded and an additional 4,000 killed. His findings were widely reported on television news programs by CNN. Such losses fortunately failed to materialize. Why was Epstein so wrong? Was his mathematical model defective?

Epstein's analysis was based on the following crucial assumptions, all of which varied from the actual Gulf War:

- A short preparatory air war of one to three days;
- A direct, frontal Coalition assault on dug-in Iraqi troops;
- Healthy, supplied Iraqi troops motivated to resist.

This illustrates how assumptions and inputs can determine analytical results. Epstein assumed a short air campaign; we executed a long one. Epstein assumed a frontal attack; we executed a flanking attack. The Iraqi troops were demoralized, poorly supplied, and sick. However, the U.S. Central Command planning staff, with a vastly more complex model, obtained similar results using like assumptions. These results were one reason why the frontal attack course of action was rejected. Without actually examining his model, in light of the Central Command's results, we should suspect that the problem was in Epstein's assumptions about the campaign plan rather than his model. His model probably responded accurately to the implications of those incorrect assumptions.

Thus, we should always look carefully at the inputs when trying to understand why a model is producing a particular result. This may seem obvious, but too often the model itself receives the blame when results deviate from what is expected. Of course, there are defective models, but they are much less common than flawed assumptions and data, errors in other inputs, or mistaken theories of combat.

Exercises and Experiments

Exercises and experiments are the oldest forms of dynamic force-on-force analysis, and, at the same time, the area of most rapid development in the last few years. Exercises are performed to instill training and assess current operational concepts, tactics, procedures, and unit or crew proficiency. Experiments emphasize new concepts, tactics, and weapons, and explore possibilities for how they may be used.

Instrumented ranges are used in both exercises and experiments to increase the reliability of measurements. The various instrumented ranges for land, air, and maritime forces enable us to come closer to creating and measuring the conditions of real warfare than has ever been possible. The strength of using these ranges is that they allow us to measure criteria as close to real combat as possible. Even so, these methods fall short of the real thing in possibly critical ways. For example, most of the participants are not in fear of dying when they participate in these exercises, so we cannot capture all of the psychological dimensions of combat. Also, some important aspects of ground combat are not well-represented, such as the effects of artillery fire and air-to-ground interactions.

More importantly, exercises and experiments are elaborate and expensive, so it is difficult to repeat trials to assess alternatives. For these reasons, particular exercises and experiments tend to be one-time only events (although they may be repeated annually), and their outcome's overall reliability is low. Again we see the trade between validity and reliability in force-on-force analysis. To date, because of service preferences, the instrumented ranges have been used mainly for training rather than experiments, but this is slowly changing to explore the Revolution in Military Affairs.

Wargames

Similar to the field exercise is the wargame. Wargames generally involve less structure and more free play than exercises; there are many decision points and many branches and sequels from each choice. The annual Global War Game at the U.S. Naval War College is a good example. Players come from the Joint Staff, the services, the unified commands, defense agencies, and other government departments to examine contemporary policy and force planning issues. Wargames are like exercises in that people are directly involved, but, unlike exercises, military forces are represented abstractly and the imaginary forces operate according to game rules. Sometimes humans apply the rules; sometimes a computer performs this function.

Wargames confront individuals with a problem (for our purposes, a force-on-force problem) and require them to make decisions to solve it. Sometimes the game stops there, sometimes the players have to implement their decisions and see what would happen—at least according to the game's rules. Wargames provide practice for commanders and staffs who will actually have to make decisions like the ones modeled. Wargames help us develop a sense of how decision makers will react to a problem when they lack previous experience dealing with it.

The strength of wargames, like that of exercises and experiments, can be their high level of validity and realism when they are done well. These methods can convincingly expose individuals to situations and conditions they are unlikely to experience before they have to confront the "real thing." They are also valuable for developing new ideas and courses of action that, in turn, need further exploration. Unfortunately, wargaming's major weakness is an appearance of reality that can frequently give participants the sense that they have encountered something close to reality without really having done so; they do not faithfully replicate the real world. People come away from such experiences feeling that they learned something when they have not—at least not about the real world.

Wargames are usually too elaborate, expensive, and time-consuming to permit repeated analytical trials to test alternatives under a variety of different assumptions and conditions. Even if resources were available, wargames, like exercises and experiments, have so many decision points that it is virtually impossible to duplicate the results of a game. This is an inherent reliability problem. Nearly every juncture in a wargame involves some sort of decision, which, in turn, prescribes the path of the game while eliminating future choices. Large games involve literally millions of these decisions. Replaying the game, changing only one decision, is impossible. Thus, it is completely inappropriate to conclude from a wargame that "A" caused "B" or that the outcome of the game reliably forecasts the result of a combat situation.

In spite of these issues, wargaming is quite useful, as long as we keep the results in perspective. Wargaming can help us train participants to think through a situation well before it or something similar occurs. Wargaming provides some insights into the broad trends that might be present in a potential engagement and that deserve further analysis. For example, a wargame might reveal logistics bottlenecks, an imbalance of air power, or a real advantage if armor is used in a certain fashion. We would then seek to assess with other methods whether these findings can be substantiated or whether they are an artifact of the wargame's rules, scenario, assumptions, and participants.

We can conclude very little based only on the findings of a single wargame. Wargames, therefore, should be used primarily for training and developing hypotheses for subsequent analysis. One application of wargames, which has become increasingly common, ignores this limitation. These wargames are large, involve high-ranking personages, and receive much advance publicity. They are held to explore high-level controversial questions of intense interest to the sponsor. Some recent examples include games about Revolution in Military Affairs-type forces and systems, information warfare, and the effects of modern air power.

The problem with these wargames is that they are too high-profile to permit open acknowledgment of their weaknesses. In addition, the sponsor too often already has the conclusions that he or she seeks to prove through gaming. The pressure is too great on the participants to produce "meaningful results," often at the last minute, which then may become institutionalized. In fact, these wargames almost never produce analytically justifiable results. They do, however, serve the political purpose of giving a group of influential people a sense of ownership over a policy for which the sponsor seeks support. It is vital that we understand this distinction when we prepare for a wargame—understanding our decision maker's objective is seldom more important than during a high-level political game.

Future Issues

We noted that most existing force-on-force analysis models depend a great deal on the combat data collected from previous modern wars, particularly World War II and the Korean War. We know that technology and economics are changing the nature of military operations rapidly. Has warfare become so different today that the data from wars fought 50 years ago is almost totally irrelevant? Developments in sensor, information, materials, and communications technologies may make it possible to conduct operations in ways radically different from the past. As a result, some argue that the current generation of force-on-force models is incapable of representing the implications of the Revolution in Military Affairs and that we need a new generation of force-on-force analysis models that are not rooted in the past. In a similar vein, there are virtually no models for analyzing peacekeeping and peacemaking operations, humanitarian assistance, and similar military activities intended to shape the security environment. This was not a great problem when the primary scenario for force planning was a Soviet attack on Central Europe. Today, these operations have become the norm and we expect them to remain so. Therefore, we need new analytic methods to help us under-stand how to plan forces with these operations in mind.

The U.S. defense community is reacting to these problems in several ways. First, the Joint Staff has commissioned a new set of force-on-force models under the aegis of the Joint Analytic Model Improvement Program. The centerpiece of this effort is the development of a state-of-the-art joint, campaign-level model called the Joint Warfare System or JWARS. Scheduled for completion in 2002, JWARS is intended to address many of the difficulties assessing issues concerning the Revolution in Military Affairs. It will represent concepts such as deep maneuver; the sophisticated use of air power; the effects of advanced command, control, and communications; special operations; weapons of mass destruction; advanced logistics concepts; and missile defense.

Second, all four services and the Joint Staff have embarked on a program of field experimentation of new technologies and operational concepts. These include the Marines' SEA DRAGON; the Army's use of the National Training Center and its high technology Force XXI (a division-sized unit); the Navy's Fleet Battle Experiments; and the Air Force's Joint Expeditionary Force Experiment. In addition, the Secretary of Defense has designated the U.S. Joint Forces Command as the principal designer and integrator for an aggressive program of joint force experiments.

Third, the services have developed sophisticated battle laboratories that consist of highly detailed models and simulations to assess problems specific to each service. The battle labs for the Army, Marines, and U.S. Special Operations Command are the focal points for DoD's analysis of peacemaking and peacekeeping. The Air Force battle lab is the focal point for assessing advanced air concepts. The Navy Warfare Development Command, co-located with the U.S. Naval War College, is the Navy's clearinghouse for innovation, doctrine, new warfighting concepts, and organizing experiments to test new tactics and procedures with the numbered fleets.

JOINT WARRIOR INTEROPERABILITY DEMONSTRATIONS[6]

To test new ideas with operators and encourage the services to accelerate their use of the most promising emerging technologies, the Joint Staff annually sponsors a set of demonstrations called Joint Warrior Interoperability Demonstrations (JWIDs). Beginning in 1995, government and industry joined forces in JWIDs to demonstrate new and emerging technologies that will shape the battlefield of the future. The projects introduce off-the-shelf, new, and evolving technologies that solve command and control, communications, computer, intelligence, surveillance and reconnaissance interoperability issues for joint and combined warfighters.

A JWID is carried out over two years with an annual event each summer. Calendar Year 2000 was a Theme Year, and 2001's Exploitation Year has followed it. In the Theme Year, participants competitively assess technologies from the private sector in a military environment. The individual sponsoring combatant command for each technology demonstration and the Joint Staff establish technical criteria and specify the goals the demonstrations must achieve. In the following Exploitation Year, the "best of the best" from the Theme Year are more fully developed into an integrated evaluation. DoD, the CINCs, and the services can target these "Gold Nugget" technologies for rapid prototyping or fast-track acquisition to speed their integration into Defense Department systems.

A different military organization runs each cycle. U. S. Space Command (USSPACECOM) hosted JWID 2000 over three weeks in the summer of 2000. The JWID 2000 theme was space-based support to warfighters: integration of space forces and space-derived information with air, land, and sea forces. The demonstrations showcased global dominant battlespace awareness in combined and coalition task force settings, and the ability to unify, integrate, and expedite intelligence, surveillance, and reconnaissance support to the warfighter through a single interface. Participants also evaluated enhanced information superiority technologies in a multinational environment.

In addition to activity at their headquarters at Peterson Air Force Base, Colorado, USSPACECOM supported other JWID 2000 warfighting commands, including U.S. Pacific Command and U.S. Joint Forces Command. Numerous North Atlantic Treaty Organization nations, Australia, and New Zealand ran their own demonstrations based on scenario inputs and included command and control interoperability trials with the United States during JWIDS.

With JWIDs, DoD is creating a process that combines the operator's experience and requirements with the practical knowledge of industry and the science of the laboratory. JWIDs put low-cost, low-risk leading-edge technology in the warfighter's hands as expeditiously as possible. Also as a result of JWIDs, DoD identifies potential investment strategies toward long-range solutions to integrate programs into an enduring, interoperable system of systems.

6. More information on JWIDs can be found at website <http://www.jwid.js.mil>.

Summary

The necessity of doing force-on-force analysis places us in the difficult position of choosing among partially satisfactory alternatives. We cannot go to war as an analytical exercise, and we certainly cannot repeat that war to test different assumptions, systems, and forces. Instead, we have to find analytical proxies for real wars. Given the complexity of the real thing, we should not be surprised that, as we have seen in this chapter, each proxy for war so far developed is unsatisfactory in some way. Figure 8-14 summarizes the trade-offs we make between validity and reliability as we choose between models. While computers have helped us improve reliability, often dramatically, they have not resolved the fundamental uncertainties associated with our competing theories of combat.

Yet, there is no alternative to force-on-force analysis if we seek to plan military forces rationally. Decision making by procedure is entirely inadequate for planning future forces in the face of rapid changes in technology and the security environment. We are left with the art and science of mixing experience and analysis as best we can to compensate for the weaknesses of both. We value static models for their high reliability – their simplicity and their clarity. Indeed, the endless spreadsheets used in dynamic model databases are themselves static force-on-force models.

Dynamic modeling can replicate actual combat better than static models. Additionally, dynamic methods can give us powerful insights into how new systems and concepts will perform in combat. However, we must view the results of such analyses critically on the basis of experience. Whenever we encounter results that defy experience, we must inspect them in depth. To do so, we may use one type of force-on-force analysis to strengthen another. For example, results obtained from a mathematical model can be tested in a field exercise, or we can use exercise results as the basis for inputs into a mathematical model. The results of either type of analysis can be used to modify static measures. Ultimately, we must be patient and thorough, understand the strengths and weaknesses of each analytic method we encounter, and resist the impulse to surrender to the frustration from using necessarily flawed tools.

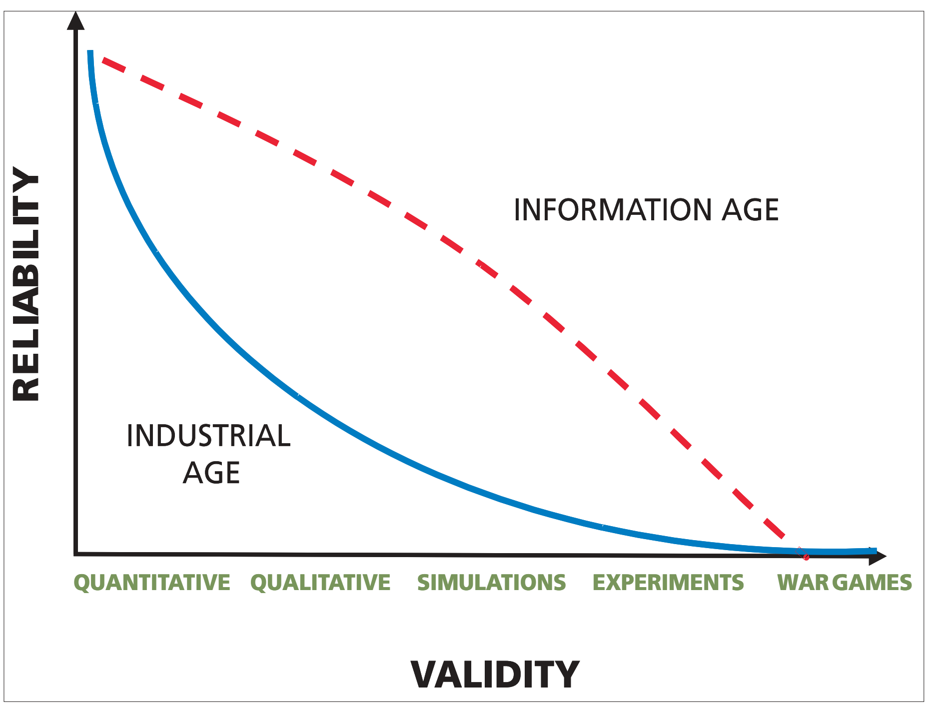
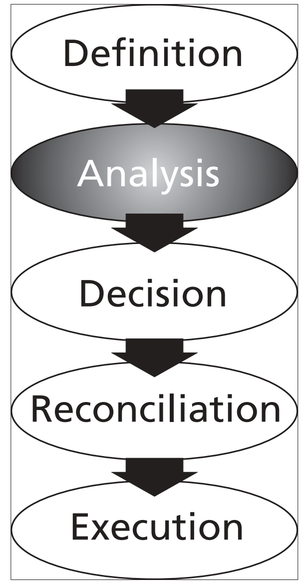


Figure 8-14. Validity and Reliability of Force-on-Force Models.

C H A P T E R 9

POLICY ANALYSIS



Find out the cause of this effect, or rather say, the cause of this defect, for this effect defective comes by cause.
-William Shakespeare, *Hamlet*, Act II, 1601

Our discussion of analysis in the previous chapters emphasized the comparison of concrete, tangible, and measurable alternatives and courses of action: Is system A more cost-effective than system B? Do we prefer this deep-water port to the alternatives? Does this Operational Plan offer greater chances of success than the others do?

But sometimes we must evaluate problems that are less clearly defined, less tangible, and less easily measured. For example, what will improve quality of life more: a pay raise or better medical care? What is the most cost-effective source of officers? Should basic training be gender-integrated or gender-segregated? Should the commissary sell discounted cigarettes? Policy analysis assists the decision maker as he or she establishes guiding principles and rules concerning social, welfare, equal opportunity, medical issues and the like.

We conduct policy analysis in one of two general circumstances. First, after a policy decision is made and implemented, we do a Policy Evaluation to assess whether the state we sought was achieved. Policy evaluations are generally descriptive in nature; they look backward to measure previous performance and compare it to what we expected. The second type is the Policy Recommendation where we identify the courses of action most likely to produce a favored outcome. Because we are recommending a future course of action, policy recommendations are prescriptive and normative in nature; they seek to solve problems and tell us what we should do and how to proceed. Occasionally, we may integrate the two types of policy analysis, e.g., a task force studying recruiting shortfalls may also recommend policy changes to improve recruiting.

Policy analysis is based on the same logic as analysis of any other question. In practice, however, we may have to take special steps to deal with the elusive character of some policy questions, which means taking care to cope with the special issues that arise with policy analysis.

Values and Policy Analysis

In policy analysis, many difficult problems are defined in terms of values. It is frequently impossible to develop criteria for evaluating policy alternatives without involving questions of values and norms. Quality of life problems are a good example. How should we measure quality of life? Is the divorce rate a good criterion? What about community involvement? Net income? How would you weight these criteria? We quickly get into problems that are both philosophically and analytically difficult. Even if our values are not engaged, others' will be. Contrast this with the analytic situations we discussed earlier. Choices between aircraft or deepwater ports can involve strong views but seldom values and norms.

Certainly, values and norms are not as central to selecting a port as they are when we make choices about equal opportunity policies. Groups may define the same problem differently from their ethical perspectives, which is why we must emphasize clarity in the Definition Phase. Ethical values may affect our measurements and data collection, intentionally or otherwise. By carefully distinguishing facts from values,[1] we can separate the ethical issues from the still-valuable tools of problem structuring, mathematics, and modeling, which should not be controversial.

What difference does it make that policy analysis is often value-laden? It need not matter much if we keep some points in mind. First, we believe that subjective does not equate to irrational. Values may be subjective, but that does not mean we cannot deal with them as objectively and analytically as we do other issues. Second, we should be intellectually honest and discriminate scrupulously between value judgments and the factual aspects of the analysis.

Cause and Effect

Another issue of particular importance in policy analysis is establishing causality—understanding why something happened. Understanding the causes of a policy problem, like low retention, is usually a precondition to evaluating alternatives for remedying it. Of course, establishing cause and effect is important in all forms of analysis, but it is especially difficult in the analysis of policy.

This difficulty exists because the world of human interactions is especially complex and difficult to understand. The scientific method is difficult to apply while evaluating policy alternatives; we seldom have a controlled environment for policy experiments. Instead, we are compelled to perform what are called natural experiments: determining cause and effect in the real world as people live their lives. This poses serious challenges to good analysis. Interventions into people's lives on the basis of a faulty understanding of social causes and effects can create serious problems. Misunderstanding the causes of poor retention can lead us to expend valuable resources while the problem continues unabated.

Because we face serious challenges isolating cause and effect in policy analysis, the Definition Phase is especially important. We must carefully define the effect we are studying. If we cannot measure the effect directly, we may have to measure it by proxy. For example, how do we measure morale or unit cohesion? Re-enlistment rates, numbers of disciplinary incidents, and field exercise scores are all measurable indicators of a unit's health. Taken together, and perhaps with a few more added, they reflect the more intangible elements of morale and cohesion.

After we identify the effect we are analyzing, we build a list of possible causes. The range of possible causes of a given effect may be tremendous. The range of plausible causes should be smaller, but may still be quite large. We use three general guidelines for linking cause and effect:

1. The extreme, complete isolation of values from decision making is called Scientific Instrumentalism, which is not what we are advocating.

- The cause must precede the effect. A must come before B.
- The relationship between the proposed cause and effect must be plausible:
  - *Sufficiency*: A must be able to cause B.
  - *Quantity*: There must have been enough of A to cause B.
  - *Duration*: A must be present long enough to cause B.
- We must account for all other possible causes.

The requirement that causes must precede effects is obvious but sometimes (surprisingly) overlooked. The plausibility requirement is also straightforward in theory but may require careful examination. The most difficult guideline we assess is the third, accounting for other possible causes of the effect.

The first category of other possible causes we examine is the group of events that would occur whether or not a particular policy is implemented. For instance, we may change an aviation squadron's training syllabus to improve aircraft carrier landing performance. While this change may contribute to better performance, we also know historically that as squadrons progress through the training cycle, their carrier landing grades improve with repetition regardless of the training syllabus. We need to determine if the squadron's scores improved beyond the historical average and, if they did, whether our change was the proximate cause. The decline in health among a graduating class tracked since it began service may be due to aging, not a degradation in the quality of military medicine.

The underlying trends of a particular environment may mask the impact of a new policy. To attract high school graduates to join the services, recruiters continually change their advertisements and selling points. A change in the number of high school graduates enlisting may be due more to the overall number of students completing high school and the strength of the economy than as a result of the recruiters' marketing plans.

Singular or short-term events may also produce outcomes that complicate cause and effect relationships. The popular film *Top Gun* led to a surge of interest in Naval Aviation. The alleged misbehavior and sexual harassment by and of some naval aviators at the 1991 Tailhook Symposium in Las Vegas had the opposite effect. In either case, recruiters and analysts must be aware of the influence of these events when they are assessing the success of their recruiting programs.

As we evaluate the effect of a policy, the other contributing causes may confuse us in several ways. At one extreme, we may decide a policy is ineffective because we cannot detect an impact. In truth, the policy may hold the line against further deterioration; its beneficial effects are negated by other causes and we wrongly classify it as unsuccessful. The other extreme is where the outside influences create the appearance of success and the policy is not actually influential. In between these poles, we have outcomes that are distorted, for better or worse, by causes apart from our policy.

We account for the contributions of these additional factors as accurately as possible. The most common method we use is creating control groups to measure the state of a population we did not expose to the policy. Complications from other causes may force us to exercise judgment and make estimates or value-based assumptions to continue the analysis. As before, we should display these assumptions and judgments clearly before presenting the results of our analysis.

Once we identify the plausible causes of an effect, we focus on the actionable causes. These become the basis for our policy alternatives and forecasting outcomes. Typically, we compare policy alternatives in cost-effectiveness studies. Unlike other forms of analysis, disparate stakeholders are more likely to value the same policy effect differently; therefore others may challenge our evaluation of the worth of an outcome or benefit later in the process. Also, we may have to evaluate several outcomes from each policy option, further complicating our comparison of alternatives and their merits.

The most difficult circumstance we face is one in which we cannot isolate cause with reasonable certainty because too many other confounding factors complicate our measurements. We may not be able to isolate the contribution of the policy we are studying. While we may dispute the cause, at the same time the present state may be clear. Polarized groups may use the same facts and the same analysis to support opposite arguments about their cause. Consider the difficulty we have identifying cause and effect if we are evaluating the military's participation in illegal narcotics shipment interdiction. Has military intervention caused any reduction in drug traffic or has the traffic been re-routed? Is the street price of narcotics a more reliable measure of effectiveness than the amount of drugs captured? In these circumstances, the principal contribution of policy analysis is to clarify the facts and structure the problem for a rational debate.

Spillover Effects

Just as there may be outside confounding causes for an effect, our policy may have consequences beyond the objective we are trying to attain. The spillover effects may be more significant than our original policy objective. For example, while managing manpower reductions after the Cold War, the Navy exempted officers from Selected Early Retirement Board actions if they requested retirement within two years. The Navy leadership was trying to humanely manage its reduction of senior officers who were likely to be separated involuntarily by offering them stability at the end of their careers. Many officers submitted two-year retirement letters, many more than the Navy expected. On the other hand, officers who intended to remain in the Navy, especially those selected to assume a major command, did not tender retirement requests. Unfortunately, so many officers requested retirement that the Navy, forced to reach its Fiscal Year end-strength levels, was forced to attrite officers who were on its Major Command list. Navy leadership could have avoided this unintended consequence with more thorough policy analysis.

Policy Analysis Methods

Next, we will examine the general approaches to policy analysis, the important role of forecasting, several of the most common modeling techniques, and some concerns about data collection to support policy analysis.

APPROACHES TO POLICY ANALYSIS

There are three principal approaches we use to analyze policy: the Top-Down approach, the Bottom-Up approach, and Mixed approach. The first two represent extreme views and the third is a compromise between them. The top-down approach is essentially the straightforward application of rational cost-benefit analysis to policy problems. We define the problem cleanly and exhaustively, we develop alternatives that represent the entire range of possibilities, and we compare those alternatives using valid criteria and a model to forecast all the possible outcomes of each policy option. Because it requires an all-powerful leader who makes decisions based on experience and analysis and who can direct implementation by an act of will over an entire organization, it is also called the Rational Comprehensive Approach. Secretary of Defense McNamara's implementation of the Planning, Programming, and Budgeting System and the Department of Defense reorganization Congress directed in the Goldwater-Nichols Act of 1986 are historical examples of this approach.

There are two problems with this approach. The first is that we can almost never meet the prerequisites of the pure top-down approach: a clearly defined problem, a full range of alternatives, easily measured criteria and a good model of the policy problem. Usually, when we think we can meet these requirements, we are proven wrong. Note the chronic under-performance of centrally decided policies, e.g., planned national economies and welfare programs. Second, even if we felt we had these requirements in hand, the chances are great that given the subjective dimension of policy questions others would strongly disagree. Seldom can we, or even very powerful decision makers, implement exactly the policy option that looked best in a formal policy analysis. Typically, decision makers have to implement a somewhat altered option, in response to other factors.

The bottom-up approach represents the other end of the spectrum: a grass roots approach to policy analysis. It responds to criticisms of the top-down approach by taking a humble view of what is possible. We accept that many policy problems must be ill-defined, that we cannot be confident that we know the full range of alternatives, that criteria for assessing their performance will be crude, and that our ability to forecast policy outcomes is poor. For this reason, we avoid defining our long-term objectives in detailed or actionable terms. We focus instead on small, achievable objectives that do not require wrenching changes from the status quo. Because the scope of change is small, our demands for forecasting are short term and easier to satisfy. Because different groups within an organization may make decisions independently that affect the whole, and the overall direction of an organization may be unplanned—or self-synchronized—this approach is also known as Disjointed Incrementalism. Secretary of Defense Cohen's decision to allow each service to decide independently whether to gender-integrate its basic training is an example of the bottom-up approach to policy.

Using the bottom-up approach, we adjust our alternatives and our objectives simultaneously as new facts become available; it is an ongoing process where means and ends merge. Because we are making decisions in small increments, there is no single large decision point; rather, we weave through a network toward the ultimate objective. The incremental effect is a continuous, responsive application of policy to solve problems rather than dramatic, isolated events. This incrementalism encourages us to involve many disparate groups to formulate policy, thereby gaining some comprehensiveness, and facilitates reconciliation because the stakes of any individual decision are never very high. If participants in a decision cannot reach consensus, we may transition to an adjudicative process with a common superior (always available in DoD) before we proceed further. Because it facilitates consensus solutions, the bottom-up approach is further known as the Political Approach.

We recognize the bottom-up approach has serious limitations, too. It is conservative in nature because it favors only small changes from the status quo; therefore it may discourage innovation and creativity, sometimes to the point of immobility. The bottom-up approach does not lead to rapid, sweeping change, though it may incrementally produce great change over time. This means that urgent problems are likely to fester which may not be acceptable. Also, the bottom-up approach means those problems of injustice, abuse, mistreatment, and the like may be changed only gradually. Congress has formed many panels and commissions to evaluate national military strategy and the services' force structures, roles, and missions, most recently the Commission on Roles and Missions (1994) and the National Defense Panel (1996), and now mandates the Quadrennial Defense Review (1997, 2001) because it considers the Pentagon locked into incremental approaches to national security issues despite the major changes in the global security environment. Because we avoid focusing on a long-term objective, our weaving trail of small steps may lead to an inefficient or even circular path.

Obviously, we will take a Mixed Approach in practice, since both extremes have serious difficulties. In the Definition Phase we recognize that the character of the problem and the conditions in which it occurs must drive the method chosen to attack it. But we have a preference for the top-down end of the continuum when we have a good grasp of a policy problem. When we do, the top-down approach enables us to exploit that knowledge for the benefit of our entire organization in DoD, which we know is responsive to positive leadership.

By its nature, the bottom-up approach says the best we can do is muddle through. That may be true some of the time, even most of the time, but not all of the time. We recognize that our ability to understand policy problems is limited, but it is not nonexistent. Some problems we can understand better than others. For example, we can understand the disastrous impact of prejudice in promotion policies. We may have some knowledge of what improves retention and quality of life. On the other hand, our knowledge of the long-term impact of gender-integrated versus gender-segregated basic training may be more limited. The kind of policy analysis that we can do for these problems is necessarily different.

FORECASTING TECHNIQUES

Prediction, as with analytic models, is very important in policy analysis. While the methods are simple, we have a difficult task judging when they are valid for the issue we are studying. We classify forecasts in three groups: forecasts by extrapolation, forecasts by theory, and forecasts by judgment. Forecasts by extrapolation project a trend into the future, usually by extending the curve or line on a graph as in figure 9-1. Here, the population's level of knowledge about Shipboard Electrical Safety is a function of how often the unit holds training; the level of knowledge of a ship's crew is highest right after a training event. As time from the training event passes, knowledge among the respondents decays exponentially until it reaches an unacceptable level unless the unit conducts additional training.

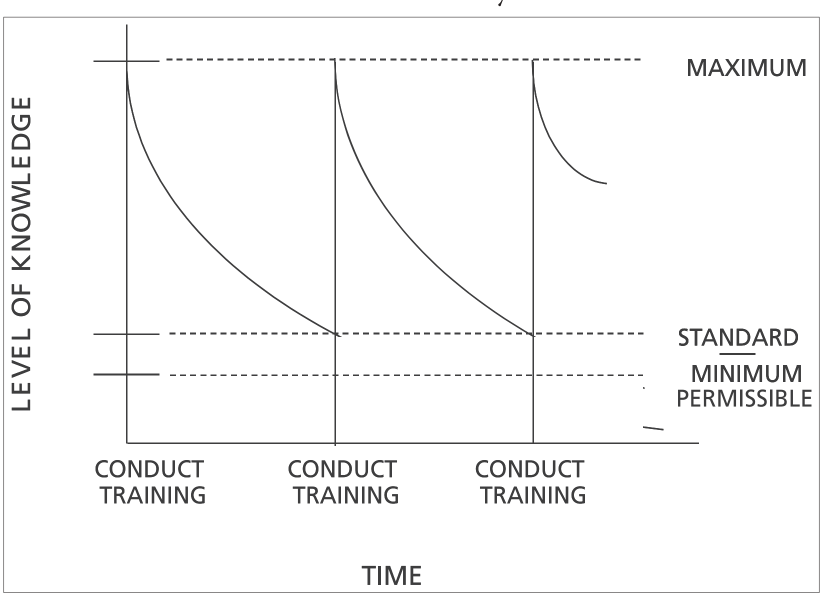


Figure 9-1. Decay of Knowledge Over Time.

Because personal safety is the issue, the prudent commanding officer holds training well before knowledge has decayed to the minimum permissible level.

The strength of this approach is its practicality, but its validity may be a problem. First, extrapolation is appropriate only when we have some reason to believe that the future will resemble the past. It often does, but it often does not. Certainly extrapolation cannot be used to predict revolutionary change that, by definition, differs considerably from the past. Second, the further we project behavior into the future, the less accurate the forecast becomes.

Theoretical forecasts require a well-defined statement of cause-and-effect concerning the issue we are assessing. That theoretical statement is the basis for the forecast. The strength of this approach is that its predictions do not depend upon straightforward extrapolation from the past. In principle, therefore, forecasts based on theory have a better chance of anticipating sharp and discontinuous change. Its weakness is that it requires a very good theory, which is often a resource in short supply. An example of forecasting by theory is Input-Output Analysis, a matrix display we can use to identify how a group of events interacts with one another. We list events or outcomes along the X- and Y- axis of a matrix and evaluate, in X and Y pairs, the effect one event is likely to exert on the other. For example, we may have a series of policy choices on one axis and outcomes on the other: How do changes in promotion policy affect retention, billet seniority profiles for the force, etc.?

EVENTS	COLLEGE BENEFITS	MEDICAL BENEFITS	COMPENSATION	DEPLOYMENT FREQUENCY
RECRUIT	High	Low	High	Low
COMPLETE FIRST ENLISTMENT	Medium	Low	Medium	Medium
1ST RE-ENLISTMENT	Low	Low	High	High
2ND RE-ENLISTMENT	Low	Medium	High	High
CONTINUE UNTIL RETIRED	Low	High	Medium	High

Table 9-1. Cross-Impact Analysis of Events and Incentives.

The Cross-Impact Analysis is a specific type of input-output analysis that shows us three aspects of linkage: whether an event positively influences another, how strong that influence is, and how that linkage behaves over time. Table 9-1 is a cross-impact analysis reflecting one analyst's impression of how certain benefits affect an individual's decision to continue his or her military career. From the table, we see that as he or she progresses through a career, college benefits become less important to the average service member and quality of life (deployment frequency) becomes more important.

We often use this kind of input-output analysis in DoD to evaluate issues like logistics support and base closures, e.g., if we close a base, how many dependents and retirees will be affected by eliminating the commissary, exchange, and medical services?

Forecasts by judgment are based on expert opinion about the future of the issue under study. As such they are inherently subjective and value-based, but, in many cases, they may be the best we can do. The Delphi Method we discussed in Chapter 5 is most common way DoD uses to make forecasts by judgment.

We may also use the Delphi method to produce diverging, not converging, viewpoints when we desire a range of opinions. After the initial round of statements and propositions, we select members to debate their positions before the rest of the group to sway opinions or influence values. The moderators summarize positions in a way that emphasizes the polarity in responses, not consensus. The final report contains a variety of options and arguments for and against each alternative. We seldom generate actionable recommendations from either Delphi process, but the forecasts they provide are very useful for policy discussions and further analysis.

POLICY MODELS

We can describe policy-modeling techniques in the same terms of abstraction, predictive qualities, complexity, and purpose, as we did earlier in Chapter 7 to describe analytic models. Policy models tend to have a higher level of abstraction than analytic or force-on-force models. The higher level of abstraction in most policy models increases the chance that we may use the wrong problem formulation or incorrectly address cause and effect.

Because they model human behavior, policy models begin with and incorporate high levels of uncertainty. Policy models tend to be simpler than other types of models because of their dependence on abstraction and assumptions; if we make them more complex we are forced to layer assumption upon assumption and our level of uncertainty quickly rises to unacceptable levels. We must identify and explain the method of prediction used by the model. We should also be clear whether we are using the model to evaluate policy or develop policy recommendations.

DATA COLLECTION

Questionnaires and surveys are the standard tools of the policy analysts. We review them carefully before we distribute them to remove bias, maintain their neutrality, and to ensure they collect information germane to our discovery of cause and effect. Closed-ended questions (multiple choice and true/false) can influence and even predetermine responses while open-ended questions allow people to answer in their own words but are more difficult for us to process. DoD gives officers a questionnaire when they resign from the service to elicit information to support retention efforts. Do those officers give their real reasons for leaving, particularly if they desire to continue as reservists? What about asking officers who remain on active duty why they are staying?[2] Groups with biases, or agendas, may deliberately try to manipulate data collection and we must safeguard against contamination—unless their values are what we are trying to identify.

Once we are satisfied with the questionnaire, we have to examine the sampling plan. The most thorough and expensive method is a survey based on sophisticated sampling. The least dependable method occurs when we allow respondents to self-select as with postal, electronic mail, and media talk show surveys. While a small percentage of the population may feel strongly enough about a particular issue to pick up the telephone, their opinions are unlikely to reflect the feelings of the population at large. Similarly, in an effort to improve physical fitness, a unit commander may consider starting an intramural sports program on base. To determine the level of interest, he may ask for volunteers to organize a base league for a sport. If he gets a strong response, he may anticipate an improvement in the fitness level of his organization. In reality, those who enjoyed the sport already will come forth while the couch potatoes remain firmly planted. There will probably be no improvement in the level of fitness of the people he was trying to reach. The data he asked for did not help solve his problem because it did not relate to the cause to achieve the desired effect.

---

2. On each set of Navy officer orders, the Chief of Naval Personnel includes a paragraph stipulating each officer will fill out a retention questionnaire. An informal survey of 250 officers we conducted at the U.S. Naval War College determined, at most, an officer had filled out two in the course of his or her career and most officers had never filled out any.

The size and composition of the sample population we survey will be very important to the credibility of our analysis. The sampling plan should use methods like random selection and geographic distribution of the sample. The sample size must be statistically significant to represent the whole population of interest. To reduce the effects of self-selection, we can collect data from a captive audience representative of the general population by conducting site surveys. In addition to surveys, we may be able to use process measures that are already in place collecting data such as complaint filings or number of transactions. We may also be able to compare organizational performance with existing professional standards, especially when we are studying a customer service problem.

We must consider how measurement error influences our evaluation of the policy effect. Measurement errors in policy analysis are usually due to flawed survey tools (poor questionnaires) or inappropriate sampling techniques. While we generally leave data collection to the analysts, we are interested in their methodology and sampling techniques. Both can be a source of debilitating error and can fatally influence the quality of our analysis. Because policy analysis generally concerns human behavior, people are the source of our most important and difficult-to-evaluate data. Analysts or respondents to surveys may introduce measurement error.

No matter how well designed and statistically reliable our study may be, the fact that we are doing a study influences the data we collect. Analysts coined the term "Hawthorne Effect" after a study in a factory of the same name. The analysts sought to measure the effect of lighting on worker productivity. The researchers observed that any change in the quality of lighting (better or worse) increased worker productivity. The analysts attributed this unexpected result to the workers' perception that because they were singled out for observation, the company was interested in their well-being. The workers' morale increased, they had more interest in their work, and productivity increased—whenever the analysts observed them. This phenomenon is the analyst's counterpart of Heisenberg's Uncertainty Principle in the physical sciences. Heisenberg, an atomic physicist, posited we cannot measure anything without altering it or its environment and we cannot know the extent of our disruptions with certainty. Whenever we measure, we must consider the effect that the act of collecting data has on the data itself.

EVALUATING POLICY ANALYSIS

Before we make decisions based on a policy analysis, we apply the same standards of validity, reliability, and practicality we have used throughout this book. In policy analysis, we have difficulty achieving high levels of validity because of the dominant role of values and the vagaries of human behavior. We run the greatest danger of misformulating the problem, compared to other types of analysis, during policy analysis.

Once we have structured the problem, we must ensure we are addressing the right things—the effects that relate to our objectives and their actual causes. Because we often use very abstract modeling techniques, we must examine our surrogates and proxies critically to ensure they reflect the areas we are trying to measure. While conducting policy analysis, we carefully scrutinize data collection to minimize measurement error from poorly constructed surveys and inappropriate sampling techniques to achieve higher levels of reliability. We can compensate for measurement errors by using control groups to compare outcomes between populations affected by a policy and groups not exposed to it in order to improve reliability. We balance the desire for perfect knowledge we gain from querying everyone in a pure and isolated environment with the cost in time and money of doing so. We seek a rational approach to the analysis in terms of practicality, balancing the resources consumed against the knowledge gained.

CASE STUDY: AVIATOR RETENTION

In the early 1980s, the Navy and the Air Force were both concerned about their poor retention of aviators. Each service did its own policy recommendation analysis, seeking the optimal cause for producing a desired effect of better aviator retention. Both services concluded aviators were resigning due in large part to the low pay scales in the post-Viet-Nam military—there was an especially large pay gap between officers and their civilian college graduate peers. The budget cuts of the 1970s made military career opportunities seem all the more uncertain. Both services therefore concluded increased monetary compensation would improve overall retention. For aviators in particular, they proposed paying bonuses in exchange for commitments to remain on active duty and to increase flight pay, the special supplemental hazard pay for aircrews.

Thus, both services decided improved monetary compensation would increase retention but they disagreed on the form it should take. The Navy wanted across-the-board continuation bonuses and a modest increase in flight pay to improve pay over the long term. The Navy argued that the morale of those not eligible for the bonuses targeted at specific types of pilots would suffer, creating new retention problems. The Air Force argued that bonuses might cause elitism, create pay inversions between junior and senior officers, and foster other manpower management problems. The Air Force favored a large increase in flight pay and small bonuses.

Congress authorized a program similar to the Navy proposal. Both services instituted a policy of awarding continuation bonuses to aviators who obligated to remain on active duty. Subsequently, both services experienced an upsurge in pilot retention, but the bonus policy came under fire from Congress. The General Accounting Office reported that $80 million of $103 million paid in bonuses went to senior aviators and Naval Flight Officers who would have remained on active duty without bonuses, or to flyers in aircraft communities where retention was not a problem. Congress discontinued funding bonuses for a year and then re-instituted them. This time they targeted junior aviators in communities with critical shortages.

Later, government policy analysts tried to determine cause and effect. There were confounding causes that made the contribution of the retention bonuses to improved aviator retention difficult to isolate. For example, on 1 October 1980, the military received an 11.7 percent pay raise, a 25 percent flight pay increase, and the newly instituted variable housing allowance. Apart from military compensation packages, the consensus of the studies was that pilot retention was far more closely related to commercial airline hiring than continuation bonus programs. It became clear that external confounding factors were driving the effect, not bonuses or flight pay.

The Air Force experienced the hazards of the Bottom-Up approach to policy later in the 90s drawdown. Like the other services, they reduced personnel as they scaled down force structure. The Air Force, driven in part by Congressional endstrength requirements, targeted personnel reductions by looking at short-term population numbers in pay grades rather than overall force structure into the next decade. Also, by 1996, under the programs described above, they had awarded aviator retention bonuses to many captains with six to eight years of service.

To bring personnel strength down, the Air Force offered separation pay to officers, including pilots unless they had obligated service remaining from earlier retention bonuses. They involuntarily separated pilots who failed to select for promotion and exempted them from their obligated service without penalty. Separated involuntarily, these non-promotion-selected pilots kept their earlier retention bonuses, received separation pay, and a separation bonus. After these policies became clear, over a hundred Air Force aviators submitted letters to the promotion board asking to not be promoted to major. If not selected, these officers with their retention bonuses, augmented by separation bonuses, would make a very substantial amount of money during their first few years with the airlines, which were hiring again. DoD may have been able to avoid a self-inflicted wound if the Air Force had more carefully considered the long-term impact of several short-term decisions.

Now the pendulum has swung again and the Navy and Air Force are facing severe pilot shortages. The Air Force is taking a holistic approach this time. They are studying the number of pilot billets they actually require (many desk jobs are coded for pilots to ensure the aviator perspective dominates the service culture). The Air Force is exploring cooperative programs with the airlines for sharing pilot transition training costs if the airlines will delay hiring until after fourteen years of service.

Summary

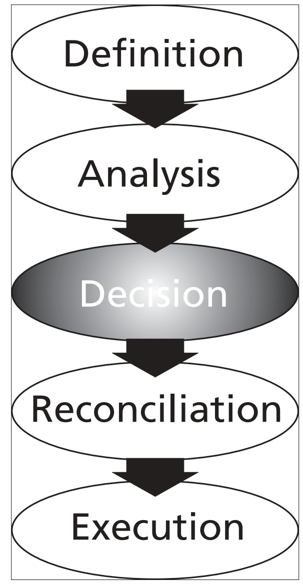
Policy analysis is the most difficult type of analysis we do in DoD. Our challenges begin with the Definition Phase because some policies are simply too vague for us to evaluate. Policy issues often require that we make value-based assumptions and consider value-based options that may not be acceptable outside our organization. Establishing the cause and effect relationship in policy evaluation or for a policy recommendation is pivotal.

The outcomes from selecting and then implementing a policy option may be varied and unintentional, therefore we make an exhaustive effort to identify spillover effects. We recommend using a mixed approach for policy analysis that emphasizes top-down or comprehensive techniques for strategic decisions (our usual situation in this course) and bottom-up or incremental approaches for routine decisions.

Because of the importance of prediction in policy analysis, we described the extrapolation, theoretical, and judgmental procedures of forecasting. Data collection and measurement errors are more likely in policy analysis because we are dealing with human behavior and responses. Good policy analysis provides information about facts, values, and actions concerning a policy issue. Because policy decisions are rooted in values and behavior, we accept that in many cases the role of analysis is limited to providing a tool to help focus debate as we prepare to make decisions.

C H A P T E R **10**

DECISION PHASE



Quick decisions are unsafe decisions.
-Sophocles, 495-406 BC, Oedipus Tyrannus

WE CAN SELDOM, IF EVER, directly implement the results of analysis without further considerations. Were it otherwise, an executive decision maker would need nothing beyond analytical skills. The ability to use analysis critically is important, but no more so than several other capabilities. Among these are your ability to understand the overall context of a problem and how your piece fits with others, the long-term paths of your and other organizations and how your decision will affect those trajectories. Also, the politics and rule sets of your organization and its ability to accept risk and take certain kinds of actions will affect which solutions are culturally acceptable. Competition for resources—the opportunity costs of adopting your preferred solution—and their effect on other decisions will influence what is possible and what is not.

All these considerations arise because our decisions and recommendations are viewed in the organizational context of changing budgets, deadlines, priorities, timetables, and bureaucratic interests. Before we can evaluate our alternatives, we must translate them so they address these aspects of the organizational environment. As we do so, we recognize there are several reasons why our analytically-produced results require additional assessment and massaging by someone with a broader perspective before we make a decision.

First, we revisit the Definition Phase and ask whether the problem that initiated our decision making still exists in its original form. What *exactly* is the problem in your view or that of the decision makers over you? We also review the Analysis Phase. Are the costs of the alternative favored by analysis acceptable to our organization's culture? Does the organization have the capability to implement the analytically-produced result? Based on your professional experience, is the analytically-produced result realistic or is it too academic, artificial, impractical, or impolitic? Do we need to implement our alternative all at once or can we break it into increments and implement it over time? We will also look ahead to the Reconciliation Phase. Will others, with different interests, embrace the solution we prefer? These non-analytical factors may make the so-called optimal alternative impractical or infeasible. We are then left to choose among alternatives that, although not analytically ideal, will still produce progress toward the organization's objective.

In this chapter, we examine the steps we take to turn the results of analysis into an implementable decision or recommendation. First, there are the important preparatory actions we take before we settle upon an option. Then, we must look at the ramifications of our choice. Finally, we prepare for the Reconciliation Phase.

The Department of Defense's Organizational Culture

The Office of the Secretary of Defense, joint staffs, defense agencies, and service headquarters all have characteristics that can make translating analysis into decisions either easier or more difficult. The DoD's bureaucratic character provides us with clarity in our organizational relationships. It values rational comparisons based upon cost and benefit, although not only those two criteria. The Department of Defense's formal structure facilitates specialization within organizations, permits advocacy and consensus among them, and provides a standing mechanism for adjudication when we cannot achieve consensus. DoD aspires to be an objective organization that rewards the best ideas and superior performance impartially and, though it inevitably falls short of this ideal, DoD comes much closer than most large organizations in this regard.

The Department of Defense is, however, an essentially conservative organization that prefers an incremental approach to problem solving. It is subject to outside political influences when making important program and policy decisions, sometimes at the expense of its own rationality. One of the premises of analytical decision making is that we have executive decision makers and organizations that are receptive to choosing courses of action based largely on their own costs and benefits. While no decision maker sets out to be irrational, other pressures may begin to crowd out his or her use of objective criteria, even on complex topics and in situations when adequate time is available for study and reflection. Alternatively, some people find decision making itself stressful and minimize their stress by making hasty decisions, small decisions, or no decision at all. Decision making is also risky, and some individuals are very risk averse. DoD has no particular exclusion from these decision-making foibles.

The art and science of executive decision making consist of giving each of these analytical and organizational factors the weight it deserves to arrive at a sound decision that is affordable, politically acceptable, and within the capabilities of our organization to absorb and implement. If the decision maker overemphasizes the analytical, he or she may select an alternative that is impractical for execution. The price of overly weighting organizational and political considerations is to choose alternatives that do not meet national security requirements and that steal precious resources from alternatives that do.

Decision Making Preparations

Good decision making is hard work. The process of using an orderly and rigorous decision-making framework and making a well-considered decision, especially in complex situations, is mentally demanding. We best prepare for making an important decision by reviewing the earlier phases of the framework to see whether anything has changed substantially since we began the process. Next, we ensure that we understand the spillover effects of selecting each alternative, and we examine the timing of our decision in the context of the problem.

OMISSIONS IN EARLIER PHASES

We framed the decision in the Definition Phase; now we review the elements we identified there to ensure nothing important has changed. We need to revalidate the problem statement and the decision objective: do we understand the decision maker's perspective or do we need more guidance? Is the problem still a problem, and is the problem statement consistent with that of the senior leadership? This is especially salient for problems that have a large political component. A political event may have triggered the need for a decision. But by the time the analysis is complete, the politics may have changed or the urgency may have dissipated. That may mean that senior leadership is now less likely to select an alternative that is a major departure from the status quo. Also, we may have formed some insights during the Analysis Phase that encourage us to adjust some aspects of the Definition Phase, i.e., we may shift some influences from external to internal or modify some problem boundaries.

In the Analysis Phase, we simplified a complex problem and applied criteria to identify the differences among alternatives and to identify our preferences among them. Necessarily, we omitted many aspects of the problem because analysis must always simplify. Now, during the Decision Phase, we need to check back to ensure that the assumptions, simplifications, and methodology we used are still appropriate for this problem and that we understand the outcomes, risk, and uncertainty of each alternative.

Thus, what we are asking at the beginning of the Decision Phase is: *Are there any major factors that we failed to address in the earlier phases that we need to consider before reaching a decision?* For example, suppose we are deciding how to reduce the costs to the Department of Defense for military family housing. One analysis used two cost criteria: member out-of-pocket expenses and cost to DoD. The least costly option to DoD, according to another analysis, is to eliminate military family housing outright and replace it with increased housing allowances that vary with location. The second study did not, however, evaluate the relative impact of this policy on different pay grades. This option will hit junior married enlisted members much harder than it will senior officers if we implement the analysis-based recommendation. The analysis does not reflect that disparity. It is not in error, it is simply incomplete. We need to stop the process and analyze some additional criteria about the consequences of each alternative on personnel before we make a decision. We also need to discuss the importance of this issue with the senior leadership.

We should note, also, whether we can combine features from different alternatives to create a new one with important advantages. Good executive decision makers seldom simply accept the alternatives as presented by the analysts.

SPILLOVER EFFECTS

As we discussed in Chapter 9, our choice of a program or policy alternative may have important side effects, intended or otherwise. An expensive weapons program may preclude purchasing or upgrading other systems (opportunity cost). Defense resource allocation decisions seldom occur in isolation.

What are the spillovers effects from each alternative? If we increase a tactical aircraft wing's training sortie rate to improve its bombing accuracy, the additional sorties affect more than aircrew schedules and air operations. The change will also affect ground operations, maintenance schedules, target range management, and logistics planning for fuel, ammunition, and parts consumption. If we cannot change the number of flight hours per month, aircrew proficiency in another area, e.g., aerial combat maneuvers or long-range navigation, will suffer. We have to think beyond our immediate expectations and consider spillover consequences, positive and negative.

Other important spillovers are organizational in nature. Every organization has limits to what it can absorb and implement. These limits may be driven by the competence of specific individuals or groups. They may be the products of important organizational traditions that clash with our choice of the analytically-optimal solution. These rule sets may cause decision makers to balk because a prominent mission or community may be diminished. Whenever our analysis recommends that we adopt a disruptive technology, antennas go up immediately to detect threats to existing organizational rice bowls; many will be more comfortable perfecting existing systems and doctrine rather than embracing change.

Organizational resistance to change, especially in the form of a disruptive technology, is nothing new. In between the world wars, the resistance to air power within the U.S. Navy by the proponents of battleships (the "Gun Club") was legendary, literally the subject of Hollywood movies.[1] The dispute whether air or surface power would dominate naval warfare continued until the Japanese attack on Pearl Harbor decided the issue in favor of air power. Led predominately by conservative battleship admirals, during the 1930s the U.S. Navy committed enormous resources to construct new battleships for the Two-Ocean Navy. The battleships, in terms of opportunity cost, were built at the expense of aircraft carriers and amphibious ships; during the war they served important but secondary roles as aircraft carrier escorts and shore bombardment platforms. But to their credit, the admirals hedged, albeit forced by Congress and a vocal minority within the service. With the huge increases in defense spending before the United States entered World War II, they also commissioned the large class of fleet aircraft carriers that came to dominate the Pacific War.

Today, DoD is considering many new, potentially disruptive technologies and organizations, e.g., unmanned combat vehicles, light armored systems, and information technology enhanced warfighter networks, at a time when we cannot fully fund both traditional and new paths simultaneously. Because every organization's culture is strongly linked to its core competencies, we may be forced to adjust or even discard a good analytical alternative if it threatens that culture too severely.

TIMING

We can make poor decisions by deciding too soon without enough information or by deciding too late, after the decision is overcome by other events. The urgency and the importance of the decision situation are often self-evident and may determine the timing of our decision by themselves; Washington deadlines are as immutable as time and tide. Otherwise-elective decisions may require our immediate attention unless we are willing to let another organization take the initiative on this issue. The nature of the decision objective influences our timing and tactics, encouraging us to act now or to delay and wait for a more opportune moment. We must understand whether a decision is time-sensitive and what the consequences of delay are. *Should we decide now?*

We may delay a decision for additional study; if so, we must specify the new information we require and decide how the delay affects the overall issue and our interests. Do we have the resources (including time) to analyze this decision further? In the early 1990s, all four services

1. For example, see 1949's *Task Force* wherein Gary Cooper and Walter Brennan lobby Navy leaders and Congress in an uphill fight to fund the first U.S. aircraft carriers and develop modern carrier aircraft before World War II.

were exploring ways to incorporate Information Warfare into their doctrines. While the other services performed a range of studies, the U.S. Air Force stood up two Information Warfare squadrons. Air Force leadership made this decision with imperfect knowledge of how they intended to employ information warfare, but these specialized exploratory units made the Air Force the pacesetter in DoD; their philosophy and systems became DoD standards that other services had to accept or work to change. This decisiveness gave the Air Force an important, tangible edge in shaping the way DoD regards information warfare.

Referring an issue to a committee or staff for further but unnecessary study to avoid a decision is a staff ploy known as "log-rolling;" it is a subterfuge for electing to do nothing. If doing nothing is our alternative of choice, then we should identify it as such. We overtly *decide* whether the time is right to make this decision now and whether to advance or halt decision making.

RESOURCES

We also need to check the validity of the resource assumptions that we incorporated into the analysis. The availability of resources of all sorts may change quickly and drastically while we are making our decision. There may also be more subtle resource issues to consider. For example, although the overall level of resources may not have changed, has the phasing of those resources been altered? Might we have less than we thought during one period of implementation and more than we thought during another? Even though we may have the anticipated resources, is there some additional reason to anticipate a change, perhaps due to some action that another organization might take in response to our decision? In recent history, DoD has been reluctantly forced to shift procurement funds into Operations and Maintenance accounts to sustain current operations; hence the lament, "procurement is the bill payer."

STRATEGY

Finally, we must keep in mind that the executive decision maker has a perspective on the problem that is broader and more informed than that of any of the analysts. That perspective is precisely what distinguishes the executive level. In addition to the other factors considered above, that broad perspective must include an understanding of the overall strategic direction that the senior defense leadership desires. An analytically optimal solution for a short-term, narrow problem may conflict with the broader, longer-term intentions of senior leaders. Staff officers may not have this information, and, therefore, the alternatives they produce may not take it into account. That comparison of the alternative to the organization's long-term goals is the executive decision maker's responsibility.

Decision Situations

Factoring in the problems of omissions, timing, spillover effects, resources, and strategy usually complicates choosing among alternatives. A cost-effectiveness analysis may produce a clear preference for one alternative over its competitors; however when we introduce organizational considerations, we may change the preference. Unfortunately, organizational factors can make it difficult to know which alternative is likely to lead to the best outcome because the definition of "best" becomes complicated. This is most likely when there are several alternatives that are roughly equal in terms of cost and effectiveness but have varied organizational implications. One alternative may be a better fit with deadlines as aging equipment is phased out. Another alternative may offer a superior fit with existing logistics pipelines while another minimizes changes to training programs. Yet another may be preferred from the standpoint of forthcoming doctrinal changes. The question is how to choose among such alternatives.

One approach is to perform a second analysis based on organizational criteria. For example, we could use a weighted model approach. A problem with this type of model is the time and difficulty to gain consensus on the weights from the collection of participating groups, each with their differing perspectives. At the very least, however, we can seek to clarify the results of selecting each alternative, even if we cannot agree on the relative importance of each outcome. Agreement on cause and effect (alternative and outcome) between organizations should be the foundation for reaching a decision about the best alternative. Below, we discuss two techniques designed for this purpose: decision mapping and decision trees. They will help us see where the commonality and differences in value of outcomes lie among our organizations.

DECISION MAPPING

Decision mapping allows us to depict the cause-and-effect linkage between an alternative (**A**) and its outcome (**O**) for each issue associated with our decision. The issues we examine may parallel our criteria and must be related to the analytic objective (and therefore the decision objective). Each alternative must generate an outcome, however some of the outcomes may be identical, e.g., where performance differences are marginal among the alternatives, we may equate the outcomes. When we look at multiple issues, the alternatives will produce different combinations of outcomes and the map becomes more complicated.

The simplest choice to map is when we have a single issue and a one-to-one correspondence between two alternatives (**A**1 and **A**2) and two different outcomes (**O**1 and **O**2). For example, with the arrow read as "yields" or "leads to," if:

Alternative 1 ⇨ **O**utcome 1 and

A2 ⇨ **O**2 and we prefer **O**2 to **O**1, we should select **A**lternative 2.

As long as we have a clear preference ordering for the outcomes for this issue, the situation remains simple even if we add additional alternatives. For example, if:

A3 ⇨ **O**3 and we prefer **O**utcome 3 to **O**2 and **O**1, then we should select **A**lternative 3.

Suppose, however, we find that we prefer **O**utcome 3 to **O**utcome 2 and that we prefer **O**utcome 2 to **O**utcome 1, but that we prefer **O**utcome 1 to **O**utcome 3. Our preferences are inconsistent. It is impossible to choose a course of action based purely on the merits of each alternative and its outcome. Within our own organization, we are less likely to face these kinds of choices, but when we deal with other organizations during the Reconciliation Phase, this circumstance becomes more likely, i.e., these organizations each have different preferred outcomes.

Another complication arises when we have alternatives that we have to evaluate for several issues and therefore each alternative has multiple outcomes. If the same alternative achieves the most preferable outcome for each issue, then it is clearly our choice. However, when they rank order differently, identifying the optimal alternative is less clear. A decision map for three issues and two alternatives looks like this:

		ISSUES	
	I	II	III
A1 ⇨	**O**I-1	**O**II-1	**O**III-1
A2 ⇨	**O**I-2	**O**II-2	**O**III-2

where **O**I-1 is **A**lternative 1's Outcome for Issue I, **O**I-2 is **A**lternative 2's **O**utcome for Issue I, etc. Now, because of the multiple issues, we may have a mixture of preferences between the alternatives as we analyze each issue. Our choice is not clear unless we can agree on the relative importance of the issues. Are they roughly equal so that the alternative that does better in any two issues is automatically preferred? Or, rather, does one issue dominate to the extent that the alternative that fares best in that area becomes our choice? Again, different organizations may value the importance of each issue differently, but we should be able to agree on the relative merits of each alternative's outcomes compared to the others' for each issue.

Consider an example in aircraft procurement when we compare three alternatives based upon various kinds of cost (usually near-term and life cycle costs), schedule (initial operational capability), and performance. We can display our preferences for outcomes in a decision map using the same notation as before. Cost, schedule, and performance are our three issues (and in this case are criteria), I through III. Each alternative aircraft has a different outcome (or value) for each issue and we show our preference for each outcome below:

		ISSUES	
	I (Cost)	II (Schedule)	III (Performance)
A1	**O**I-1	**O**II-1	**O**III-1
A2	**O**I-2	**O**II-2	**O**III-2
A3	**O**I-3	**O**II-3	**O**III-3

As we look at each issue, we will rank order our preference for the alternatives as best, median, and worst and replace our symbols. Rebuilding the map:

		Issue	
Alternative	**Cost**	**Schedule**	**Performance**
A1	**Best**	**Median**	**Median**
A2	**Worst**	**Worst**	**Worst**
A3	**Median**	**Best**	**Best**

Table 10-1. Decision Map.

Although we cannot make a quick choice between **A**1 and **A**3, we can eliminate **A**2. It scores poorly for every issue. Our analysis and mapping did not provide an unambiguous answer, but it did structure the decision and allow us to winnow the alternatives. We can now focus the discussion upon the relative importance of Issue I (Cost) compared to the combined value of Issues II (Schedule) and III (Performance) and apply our military judgment to make a decision.

Let us expand this example to make it more specific. As above, suppose we are reviewing the analysis to support a decision to choose an alternative for a tactical aircraft, an air superiority fighter. We have decided there are four principal organizational issues or criteria: Near-Term Cost, Total Ownership Cost, Date of Initial Operational Capability, and Tactical Performance. We have identified three alternatives: **A**1, an upgraded fighter; **A**2, a new fighter currently in advanced flight testing; and, **A**3, a new fighter under concept development.

Alternative 1 can be fielded soonest but it is the most expensive in the near-term. Alternative 2 has the best tactical performance but it has the highest total ownership cost. Alternative 3 has the lowest near term and total ownership costs, but it will take the longest to field and has the poorest tactical performance. The decision map looks like this:

		Issue		
Alternative	**Near-Term Cost**	**Total Ownership Cost**	**Initial Operational Capability**	**Tactical Perfor-mance**
Upgrade	**Worst**	**Median**	**Best**	**Median**
Prototype	**Median**	**Worst**	**Median**	**Best**
New design t	**Best**	**Best**	**Worst**	**Worst**

Table 10-2. Decision Map for Aircraft Alternatives.

Part of your organization prefers Alternative 2 (the prototype) because it believes that the superior performance of the aircraft will convince DoD's leadership and Congress it is worth a higher total ownership cost. Another part of your organization prefers Alternative 1 (the upgrade) believing that, despite the immediate up front cost, it is crucial to field an improved aircraft against the threat as soon as possible. Another organization, whose cooperation we need, favors Alternative 3 (the new design). They feel strongly that conserving financial resources in the short term should dominate this decision now because countering the foreseeable threat does not demand a large leap forward in capability. The three stakeholders do not share the same preferences for second choices either.

How do we make a choice in such circumstances? The first step is to see whether, although we have no consensus on a first choice, we have agreement on a last choice. In the same vein, we may seek to clarify the situation further and produce a more internally consistent set of preferences. This requires that we be thoroughly grounded in our organization's interests (see Chapter 11, Reconciliation). For example, if we bring the stakeholders together to probe their views jointly, they may converge on a choice or reject one of the alternatives. Another approach is to make a decision based purely on what we can implement least painfully. That may not be the optimal choice for anyone, but the process moves forward.

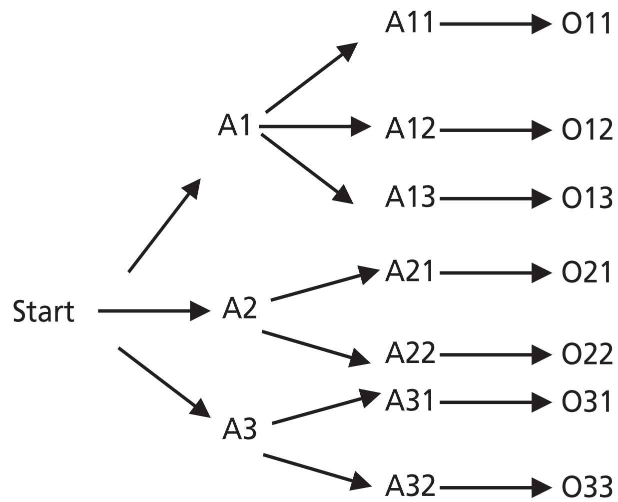
Rather than selecting the best alternative, we may choose to satisfice, settling on an alternative that solves the problem, an alternative that is *satis*factory and suf*fices*, without an exhaustive search for an optimal solution. Satisficing is rational behavior when the differences between an adequate solution and the optimal solution are small and difficult to detect. It is also rational if the decision maker is constrained by time or if his top priority is to “keep peace in the family.”

Yet another approach might be to make our choice among alternatives based exclusively on our organization's preferences and let the reconciliation process work out the differences in preferences with the other stakeholders.

DECISION TREES

We have a graphic technique, Decision Trees, that is useful for displaying chains of outcomes as they relate to a decision or analytic objective. They are particularly helpful for decisions with many spillover effects. We use decision trees in military operational planning to show the branches and sequels from our courses of action. We use them in force planning decisions, especially with policy choices, to display a series of outcomes conditioned by preceding choices. Often, as we cope with risk and uncertainty, we label them with expected values.

The decision tree begins with an initial decision point that branches into alternative paths called branches. At the end of each branch is another node (decision point) that may generate another set of branches. A sequence of branches ends when the paths reach a set of final outcomes in terms of the analytic or decision objective; the number of branches depends on the decision process we structure along the way. Immediately before the outcomes, the alternatives must be mutually exclusive.



Alternatives
A1: Aggressively punish perpetrators
A11: Discharge all offenders immediately
A12: Reprimand first offenders; discharge thereafter
A13: Warn first offenders; reprimand second offenders; discharge thereafter
A2: Educate personnel
A21: Increased training for all personnel
A22: Focused training for accessions only
A3: Adjust working environment
A31:Gender integrate more units
A32: Gender segregate more units

Outcomes: < SH means projected decreases in Sexual Harassment incidents, <<SH and <<<SH mean increasingly larger decreases.

O11:<<<SH, loss of trained personnel
O12:<<SH, fewer trained personnel lost
O13:<SH, fewest trained personnel lost

O21:<<SH, increased chill in the workplace, lost work hours
O22:<SH, less work time lost to stand-downs

O31:<SH, initial upswing in SH followed by long term decline
O32:<<<SH, probably not politically acceptable

Figure 10-1. A Decision Tree for Combating Sexual Harassment.

Decision trees are especially effective when we use them to represent an incremental process. Figure 10-1 shows the components of a hypothetical decision tree for sexual harassment policy. The decision objective is to build a policy to reduce the occurrence of sexual harassment in a command. The first set of branches shows the major approaches the command may take to combat sexual harassment: reprimand bad behavior (Punishment - **A**1), prevent inappropriate behavior (Education - **A**2), and adjust the organizational climate (Working Environment - **A**3). In this case, there are more decisions we must make about each alternative before we can establish a working policy.

To expand branch **A**1, we decide how we will discipline personnel who violate our policy; our vague intention to punish them is not enough. The choices about our baseline policy that follow are of decreasing severity and are mutually exclusive:

A11: Discharge all offenders immediately

A12: Reprimand first offenders; discharge thereafter

A13:Warn first offenders; reprimand second offenders; discharge thereafter

After we identify the alternatives, we assess the outcomes associated with each. Here we capture spillover effects as well as the outcomes related directly to the analytic or decision objective.

For the alternatives above, where <SH means we project decreases in sexual harassment incidents and <<SH and <<<SH mean increasingly larger decreases:

O11:<<<SH, loss of many trained personnel

O12: <<SH, some trained personnel lost

O13: <SH, almost no trained personnel lost

We can follow a similar process through branches **A**2 and **A**3.

Several aspects of the decision tree in figure 10-1 help us frame this decision and decide upon a policy. First, at the end of each secondary node we have mutually exclusive, actionable alternatives from which to choose. Second, the spillover effects of each choice are also clear, e.g., the penalty for drastically reducing incidences of sexual harassment (**O**11) by implementing **A**11 is a higher attrition of skilled personnel. These personnel will be costly to replace, both locally (gapped billets from unplanned losses) and for the service (recruiting and training). Note, too, that although not shown as branches on this particular tree, doing nothing is also an option for each issue: we could leave current punishment and education practice in place or decline to change the number of gender-integrated units.

Decision trees do not need to be symmetric; the number of nodes along paths in the same tree may vary and there may be any number (greater than one) of branches from a decision point. Every end branch—the actionable alternative—must result in an outcome that is acceptable in some way to the decision maker. If it is not, then the alternative leading the outcome is not valid; in this case, we need alternatives that lead to a decrease in sexual harassment. We can enhance decision trees by labeling the probability of an outcome or its expected value along the branches. Another strength of the decision tree is the way it highlights opportunities for combining alternatives. For example, if we are going to remove perpetrators with the first offense (**A**11), then we need to select **A**21 as well or the policy will not have any deterrent value and we will suffer higher than necessary personnel attrition. If we also select **A**31, we will want to postpone its implementation until we do the comprehensive preventative training or we may create more incidents and discharge more personnel than is necessary.

Reality Check

After we select an alternative from our personal and then our organization's perspective, we must take one more look at our decision using professional judgment and intuition. We have all had the experience of solving a complex math problem and arriving at an answer that, by brief inspection alone, just seems wrong. We need to put our analysis-based decisions to exactly the same test. We may have striven so hard to overcome anticipated opposition that we have lost sight of the best alternative. We may have looked at the material so long that an important issue has escaped our notice. We should revisit our expectations for the Reconciliation Phase to see if the alternative can survive politically, e.g., the Army National Guard is so politically powerful it has forestalled any serious combat force structure reductions.

Presentation of Results

We should document our decisions. Our decision presentations can range from informal conversations and memos to academic thought pieces published in professional journals. Documentation preserves *our* thoughts for easy reference, for our successors, and ourselves that may be helpful as our organization prepares for reconciliation. We can use reports and studies to record analysis-intensive efforts like procurement decisions or force structure proposals and we often distribute our results widely to other organizations. We can write issue papers for the internal consumption of our headquarters or staff. Issue papers document program evaluations and record our thoughts about the proposals of others; we use them extensively to prepare senior leaders for meetings and testimony. Policy option papers are the culmination of policy recommendations and at their conclusion we often ask a decision maker to select an alternative. Any of these formats may be the basis for decision briefings for senior leaders or information briefings for other organizations. Regardless of form, our Executive Decision-Making Framework provides a good outline for crafting any of these reports.

Preparing for Reconciliation

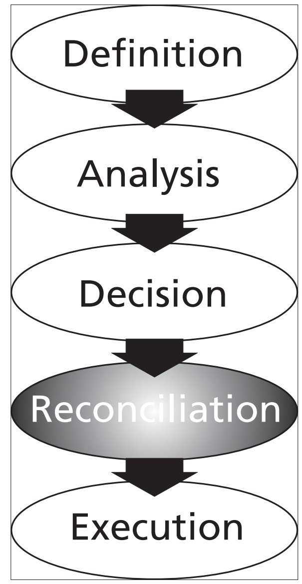
Our framework helps us make a rational choice for our organization, or at least to ask the right questions at each step in the analysis of our choice. We probably have known from the beginning that *our* choice is unlikely to be *the* choice that our command implements without the approval of other decision makers from organizations both parallel and senior to our own. We have deliberately framed the decision in terms of our organization and our internal influences to simplify the problem. As we move toward reconciliation with other groups and address external influences, the values and norms of these new participants, as they apply to the alternatives, will be very important to us.

The reconciliation of our decision with the interests of others involves advocacy and negotiation (which we discuss in the next chapter). Before we negotiate, we must be firmly grounded in our interests, our priorities, and our preferred option. We have done this by using a systematic decision process that involved the senior leaders at each important juncture. We must be prepared to accept the burden of proof if we are going to advocate a change in existing force planning options or policy; our analysis must support the change. Our advocacy must be rational itself, based on values consistent with our organization and logically supported by facts.

Summary

In this chapter, we have discussed the procedures for selecting alternatives within our organization. We began by reviewing the earlier Definition and Analysis Phases to see if anything important had changed or been omitted. We evaluated spillover effects and decided whether now is the time to decide this issue. To frame the decision and focus discussion, we examined the display techniques of mapping and decision trees. After selecting our personally preferred option, we did a reality check, and forwarded our choice to the decision maker. We documented our decision and now we are looking ahead to the Reconciliation Phase.

RECONCILIATION PHASE



You cannot shake hands with a closed fist.
-Indira Gandhi, 17 May 1982

AFTER WE HAVE DEFINED THE PROBLEM, completed the analysis and the decision maker has carefully weighed the alternatives and made a choice from the standpoint of our organization, we must consider how other stakeholders will receive our decision. Some of them may be able to modify our alternative, veto it outright, or inhibit its execution. During the Definition Phase, we identified external stakeholders and influences with an eye toward how this decision would be reconciled with those parties. During the Analysis Phase, we identified criteria that were important to our organization; now we will examine the criteria from the perspective of the other parties as well.

During the reconciliation phase, we revisit many of the choices that we or the analyst made in the course of doing the analysis. We may need to analyze criteria that are not important to our organization but are critical to selling our choice to others. The alternatives we rejected from the outset, or discarded when we chose an option, may be the preferred alternative of another stakeholder. If someone promotes a new alternative, we need to know how it affects our organization, especially if a mutual superior accepts that recommendation instead of ours. Once the decision maker approves an alternative for implementation, executing the policy or program requires that we plan for overcoming interference from competing interests and obstructionists.

Negotiation

The reconciliation process will probably involve negotiation. Negotiation is the process by which individuals work with peers, superiors, subordinates, and others to achieve agreement on issues in which they have a mutual interest but, quite often, no authority to dictate the outcome. Humans differ in their preferences, and are sometimes willing to fight for a cause or seek retribution when their positions are disregarded. Negotiations are therefore crucial for maintaining order in society at all levels, from the international community, through the state and local public forums, to the work place and home.

Additionally, negotiating skill is a hallmark of successful command leadership and effective management, because it is key to building consensus, getting individuals moving in the same direction, and accomplishing things they could not, or would not, do on their own. Those who refuse to negotiate and appeal all disagreements to higher authority are inevitably and justly labeled as poor leaders. Likewise, those who are better prepared will eventually overwhelm others who attempt to sell their position on the basis of their strong personality without concrete analysis to support their position.

Reconciliation takes many forms. In its simplest form, it involves two parties and a single issue where both sides value the outcome about equally, e.g., two squadrons competing for a use of a bombing range. Force planning decisions are usually more complicated, involving multiple parties, multiple issues, and varying preferences (and strengths of preferences) for outcomes.

In this chapter, we will introduce some concepts and strategies for reconciling our organization's alternative with other groups. We will also examine negotiation at the individual and group level, and address scenarios we are likely to encounter as senior decision-makers in the Department of Defense.

Keep in mind that the fundamental basis of successful negotiating is *analysis*, not tactics. When we clearly understand our interests, those of the other participants, and the areas in which agreement is possible, 90 percent of the negotiation is completed. On the other hand, we may be superb tacticians at the negotiating table, but without an analytical grasp of the substance of the negotiation, we are as apt to negotiate our organization to failure as to success.

PARTICIPANTS

We have two negotiation goals: to have our preferred alternative adopted and to satisfy all stakeholders' interests. We know that different organizations have different values. We may be able to achieve our goals by helping other groups achieve theirs; for instance, we may exchange concessions on objectives of little value to us for concessions on goals important to us. If the other side *thinks* they achieved their goal, whether or not they actually did, or whether or not that goal is valuable by our standards, they will be satisfied. The more we know about the other participants, the greater the likelihood we will achieve our goals through negotiation.

CULTURAL DIFFERENCES

In order to negotiate with other participants, we should examine our cultural differences and understand others' interests and their style of negotiation. For senior defense professionals, this aspect has become increasingly important as we engage in more coalition operations. There are distinct national cultural approaches toward negotiation. Americans tend to prefer a direct approach that conserves time and reaches a definite conclusion, often summarized as "time is money." Many other cultures consider trust engendered by enduring relationships as an extremely important prerequisite to serious negotiations. They will invest considerable time developing personal communication before they are comfortable discussing a serious problem. They see professional relations as open-ended and the current issue as one of many that the participants will address over the long-term. Therefore, when they say, "time is money," they mean the time spent building an atmosphere of trust and cooperation now will be rewarded later during the reconciliation of many issues.

Additionally, the seniority of the participants may affect the level of discussion during negotiations. In some societies, senior officials negotiate in terms of broad principles only and leave the details to be worked out among functionaries once they have settled upon a general agreement. Cultural differences are not based solely on nationality and religion. There are significant cultural differences in force planning among the services and between warfare and staff communities such as aviators, surface warfare officers, submariners, and supply officers.

PERSONALITIES

Negotiators sometimes have motivations and agendas that are markedly different from the position we expect from their organization. This difference may be a deliberate negotiation tactic or it may be real. First, we must ensure this does not happen between our negotiators and our organization. Second, this kind of discrepancy may present several opportunities. For example, we can "educate" the other participants about their organization's interests and thereby move toward a coalition or agreement, or we can exploit this separation. The latter course of action may lead to a tactical victory but damage the long-term relationship between our organizations. Moreover, each individual negotiator may seek personal satisfaction in addition to organizational success. Negotiators want to feel good about themselves, they want to avoid being boxed into a corner, and they want to be recognized as having done a good job. We should consider those emotions as we negotiate.

Additionally, we often deal with characters who behave badly. Nonetheless, we must learn to separate people from problems. All too often we fail to negotiate seriously with someone we do not like for personal reasons, even when there are opportunities for good solutions. This is not to say that bad behavior should be ignored or accepted, but it should be separated from the problem we are trying to solve. Aggressive or negative behavior may actually be a negotiating ploy by another participant; however, how we react, if we react, is our choice and is situation-dependent. We should gauge our reaction to how deliberate we consider the offensive activity, whether it is sexual harassment, name-calling, or obstructionism. We should also consider whether the negative behavior is due to shortsightedness or ignorance. Good ideas and proposals may come from unlikely sources; we must judge each on its merits, not its source. We should not allow some people's unpleasant nature to intimidate us into modifying the foundation of our negotiation strategy (we may, however, adjust our tactics).

INTERESTS AND POSITIONS

Before we enter into negotiations in the Reconciliation Phase, we need to identify our initial position and our interests. In our vocabulary, interests are enduring activities, rights, and concerns that are connected to our organization's mission and core values. Interests are not variable or negotiable. Positions, on the other hand, are flexible; they are our place or standing relative to our interests. Positions protect or advance our interests. For example, the Air Force needs a new fighter to guarantee it can achieve air superiority in the future; therefore it is advocating procurement of the F-22 Raptor. Its interest is air superiority; its position is buying no less than 339 F-22s. Reconciliation generally requires that we form a consensus among many participants, each of whom has his own interests and begins with a position on the problem.

We can usually discern the differences between our position and the other participants' positions. Identifying their underlying interests may be much more difficult, but the more we understand their interests, the better we can determine how much room they have to adjust their positions. The assumptions and criteria the other participants used to select their own optimum solution to the problem may reveal their interests. If we can satisfy their interests using another alternative—preferably our alternative—we may convince them to change their position. This is a key concept in negotiation.

One of the most common sources of conflict is the adoption of different assumptions and criteria to evaluate the alternatives. If we can agree on *a common set of objective assumptions and criteria* with the other participants, we may eliminate many important differences about which alternative best solves the problem. Whether we can find such a set depends upon whether our interests are compatible with those of the other organizations.

Expert military judgment, our usual basis for selecting criteria, may be grounded in beliefs based on our organizational culture that are fundamentally opposed to those held by other participants. Despite these differences, we may find grounds for mutual agreement because there is nearly always some interdependence of interests among some of the parties in any complex decision. This interdependence may be created by a continuing relationship between some participants who place a higher value on maintaining comity than on any particular decision. For example, the need for harmony among the Joint Chiefs of Staff often overwhelms parochial differences among these Service Chiefs. Additionally, participants in force planning may be mutually dependent on each other for support in other forums; e.g., the unified commanders often coordinate their positions with each other before the Joint Requirements Oversight Council visits them to discuss issues.

Likewise, separating interests from positions is a necessity for resolving labor disputes on DoD installations and for defense contractors. Labor and management often have an adversarial relationship. Labor would like to be paid more and management would like to show a greater profit by cutting labor costs, but they are dependent on each other to achieve their objectives and protect their interests by ensuring the company survives. Both parties can promote their own interests if the company makes more money. Their positions differ on how profits are distributed, but if they can agree on a solution that increases the money both sides receive—higher wages for the workers *and* greater profits for the company—both sides benefit. Their positions will shift, but they should be able to reach an agreement because of the interdependence of their common interests.

PRESSURE FOR AGREEMENT

Participants in reconciliation are under pressure to reach an agreement. That pressure may lead them to make concessions in order to form a consensus. DoD is a consensus-oriented bureaucracy. Senior officers regard a leader as ineffective when his/her organization is unable to reach accord with other groups and he/she continually forces decisions upward. A contractor may be under financial pressure to complete an agreement because his creditors demand immediate payment. The military may need a critical piece of equipment in order to conduct a particularly time-sensitive mission. As we enter negotiations, we assume that we are not the only ones under pressure to reach agreement; we would especially like to know what kinds of pressures the other participants are under.

AUTHORITY

When negotiating we also desire to know in advance who has the approval authority for an agreement (for each participant) and whether that authority will be present for the negotiations. Negotiators sometimes conceal or misrepresent approval authority as a negotiating tactic; therefore we cannot assume a party has such authority by virtue of participating in the negotiation. Generally, we recommend negotiating with the highest-level individuals possible from the other organizations, because this conveys several tactical advantages to us. For instance, higher authorities are generally less conversant with the details of an issue and they are inherently disinclined to let details block an agreement. Also, senior leaders are more pressed for time and therefore motivated to reach an agreement quickly. (Of course, these factors can also work against us when contractors approach senior DoD leaders and press them for decisions without allowing input from their more knowledgeable subordinates or operators.) Moreover, higher authorities have greater latitude to make concessions and will do so to sustain long-term relationships. Senior corporation officers generally value a customer like DoD more than their middle management does.

Who else has the power or influence to affect decisions during reconciliation? The answer may have a large effect on how a negotiation develops. Senior leaders, action officers, and acquisition specialists in DoD often rely on the advice of operators to determine which alternative their organization will support. *Cui bono* (which means "to whose advantage") is the principle that those who have the greatest stake in the outcome are most likely to work hardest to influence the decision. For example, in force planning decisions, usually the soldiers, sailors, airmen, and marines have the most to gain or lose. Contractors know the Pentagon usually solicits the operators' opinions before they select criteria and that DoD values their military judgment about the problem and alternative solutions. Therefore, contractors may lobby the operators directly, hoping they will influence the decision makers to favor their alternative.

TEAM NEGOTIATIONS

Negotiating using a team from our organization has advantages and disadvantages versus sending a single participant. Each negotiation has a setting and we should determine whether the other participants are bringing teams and, if so, the general composition of those teams. We do not want large asymmetries between our team and theirs. Sometimes we may choose to send a slightly larger or more senior team than the other participants as a statement that this issue is important to our organization.

Team leaders must strike a balance between giving their representatives enough flexibility to negotiate independently, and not allowing them so much freedom that they present conflicting proposals to the other participants. Teams may include a variety of specialists to capitalize on expert knowledge. In so doing, each member can negotiate his segment of an agreement simultaneously, and the specialists are on-hand and readily available for consultations with the team leader. Team negotiations also allow a wider range of negotiation strategies. We can test an idea (a trial balloon) to get responses with the understanding that the proposal is contingent, i.e., we cannot commit to it without speaking to our colleagues first. Moreover, teaming allows members to play diverse roles. One can play the role of the tough guy, while others speak softly, gain trust, and gather information. In fact, coalitions may develop between team members and participants from other organizations' teams. Just as in battle, organization and intelligence are keys to success. We need thorough knowledge of each alternative's effect on us, and a team may be able to make this assessment quickly and effectively on site.

The disadvantages of team negotiations stem from coordinating simultaneous actions by the team members. Each member must be thoroughly grounded in our current position and know how much he or she can concede. Additionally, disruptive coalitions and disputes may develop within our team. Also, teams diminish personal accountability for everyone except the leader; in so doing success or failure is shared. Moreover, the decision hierarchy necessary in a team may slow momentum toward agreement as members feedback their individual status. Finally, because of the shared sense of responsibility, teams tend toward deadlock more than individuals.

POWER AND INFLUENCE

Power during reconciliation is the ability to coerce or direct the actions of others while influence is the ability to affect the behavior of the participants. Different combinations of power back the positions that negotiators assume. Participants can exercise power and exert influence from many bases:[1]

- Coercive Power–the holder can inflict punishment.
- Reward Power–the holder can bestow something valuable.
- Legitimate Power–rules, values, or conviction convey moral superiority.
- Referent Power–charismatic qualities cause unquestioning confidence.
- Expert Power–specialized knowledge conveys better and more impressive technical understanding.
- Representative Power–democratic delegation of power by a large group.
- Connective Power–the holder is linked to a major power broker outside the negotiation.
- Information Power–the holder has key information that is required for a successful agreement.
- Coalition or Alliance Power–several participants unite to overwhelm a single strong power or another combination of participants.

There are also more subtle sources of power and influence during negotiations; the shape of the table sometimes is actually important. Those in the middle of a room, or table, are closer to the center of the conversation than those on the fringes and therefore they have more opportunity to participate in and dominate the discussion. Holding meetings in one participant's office implies he has more influence over the discussion because the other participants came to see him. Physical size can be intimidating. Even differences in chair heights create perceptions. Although these "status" items may not be important to us personally, many negotiators are very sensitive to their surroundings and interpret them carefully; if we ignore or dismiss these details, we may inadvertently send the wrong signals.

Additionally, some participants may gain influence during a negotiation because their social status conveys a certain type of authority, e.g., clergymen assume an air of legitimate power on moral issues and officers who rose up from the ranks are afforded expert power on issues regarding enlisted personnel. Professional reputations for expertise and honesty generate power, as does the popular mandate that results from an election.

Strategies

We should plan a flexible reconciliation strategy. That means we must know, before the negotiations begin, the issues, and how important and urgent they are to the other participants and ourselves. We know which alternative we prefer and should try to find out the alternatives others are promoting. We should decide in advance how we will protect and advance our interests and how we will respond to the tactics and positions of others. We may use overlapping approaches and include various what-if branches, but we absolutely must begin with a game plan.

GENERAL APPROACHES

There are two basic philosophical approaches to negotiations. The first mode is variously described as the Zero Sum, Win-Lose, Competitive, or Traditional approach. It is a hard-nosed

1. Based on "Power and Influence" by William E. Turcotte (Newport, RI: Naval War College faculty paper, March 1997).

technique: if one participant wins, the other must lose. Typically, we view internal DoD budget negotiations this way; every dollar gained by one participant is a dollar given up by another. In a competitive negotiation, we pressure some parties to join coalitions with us and force others to change their positions more to our favor. Often, an agreement is not possible and an adjudicator decides how to allocate resources after evaluating the positions of each participant. This mode emphasizes the positions taken by each participant.

The second approach is called the Mutual Gains, Win-Win, Both-Win, Interest-based, Value-Building, or Cooperative approach. This view envisions all the reconciliation participants working together to reach a solution that satisfies each of them. Ideally, we seek to increase resources or find ways to share them, e.g., all the services share the overhead for helicopter pilot training to reduce each of their costs. The focus in this mode is not on winning or advancing our position, but on the overall outcome and meeting the interests of each participant. There are two fundamental methods for improving the outcome for all the participants: (1) add value by increasing the group's resources and (2) find asymmetric values that participants can trade with each other, i.e., one participant may concede on an issue of low importance to him in exchange for a concession by another on an issue he values highly. This is why we emphasize separating interests from positions.

For most decisions, the most successful negotiations are a mix of both approaches. People intuitively gravitate toward the first approach, but negotiators actually realize the greatest success from the mutual gains approach. For example, suppose we are procuring some specialized radios for ground forces. Our annual budget is limited but steady. The current manufacturer can supply the radios and offers them at a high price. Typically, with a zero-sum approach, our position would be to press for a lower price or threaten to open the contract to new competitors. The company would hold out as high and for as long as possible. Under a mutual gains approach, we might offer the company multi-year contracts to reduce their costs with the expectation that they will lower their price to us. We should use a mutual gains strategy wherever possible—its advantages seem obvious—but we must be prepared to shift to the competitive form if we cannot find a basis for mutual gains or if the other parties are unwilling to cooperate.

CONFIDENCE BUILDING STRATEGIES

One commonly successful negotiation strategy is to break a problem into smaller segments and solve them individually (as we considered in the Definition Phase). At the outset, all parties acknowledge that the agreements they make on each segment are contingent on a complete solution to the entire problem. Breaking a problem into parts facilitates agreement by making trade-offs easier to identify. From there, we can take either of two approaches. The first mode is to "pick the low hanging fruit" by reaching early agreements on the issues in which there is little contention. This builds mutual confidence among the negotiators and creates an atmosphere of cooperation as we approach the tougher issues. This approach is often necessary when the stakes are very high for the participants and there is great distrust between them. For instance, the negotiators that brokered the 1993 Oslo Accords between the Palestinians and the Israelis used this technique. They reached agreements about smaller issues like combined Israeli and Palestinian police patrols and incremental transfers of land to Palestinian control. They left larger issues, like the fate of Jerusalem and Palestinian independence unresolved while they waited for the smaller steps to build confidence between the parties. By breaking the problem into smaller segments, they made the risk of failure of any one part of the Oslo Accords an acceptable risk to both sides. They also stipulated that further progress toward a permanent peace was contingent upon the success of these smaller agreements.

The other approach is to work on the tougher issues first to reduce the time negotiators spend on the easier parts of the problem. The less contentious issues should fall in place quickly after the parties resolve the hard issues. If, as we proceed through the tough issues, we cannot settle a particular segment, we set it aside and work on the next part. Each resolution of a tough issue builds confidence among the negotiators and that confidence leads to greater cooperation. On the other hand, if we have irreconcilable differences with other parties, we will find out early in the process and end the negotiations without expending many resources.

Linking the segments of a negotiation allows both parties to maintain a positive demeanor during the negotiation because of contingent if-then agreements. Rather than rejecting unfavorable proposals, each participant can approve them if they are linked to significant concessions in other areas.

GAME THEORY AND THE USE OF NEGOTIATION GAMES

Game theory has been used for over half a century to help us understand why people, businesses, and governments act as they do in a variety of situations. Games help us prepare for more sophisticated negotiations by practicing negotiation under less threatening conditions. Everything from economic decisions among competitors to arms control has been subjected to analysis through game theory.

The simplest form of negotiation games is the zero sum game. These have equal but opposite payoffs for each participant for a single move or a series of moves. Whatever one player wins, the other loses. These are games of pure competition that do not foster cooperation because the collective payoff is always zero, no matter what strategy either side follows. The limited range of outcomes (I win, you lose) affects behavior, but negotiations are seldom worthwhile because of the initial conditions.

Other games, however, have payoffs that vary. The payoff depends upon how the players interact. The simplest example of this non-constant sum game is the Prisoner's Dilemma. Although simplistic, the principles it illustrates are key to understanding and succeeding at negotiations. The police apprehend Bob and Sue, two perpetrators of a crime. They separate them immediately upon their arrival at the police station. Each suspect is given a choice of confessing and implicating his or her partner, or not confessing. Should one confess while the other remains silent, the stoolie gets off with probation while the other goes to jail for 20 years. If both confess, they both go to prison for ten years each. If both keep silent, they will each receive a one-year term for lesser offenses. In game theory format, the payoffs for each combination of decisions will look like the following matrix, in which the payoff (years sentenced—*low numbers are good!*) for Sue is always the *first* number, and that for Bob is the *second*.

Payoff [=] Sue, Bob	Bob confesses	Bob doesn't confess
Sue confesses	10 years, 10 years	Probation, 20 years
Sue doesn't confess	20 years, probation	One year, one year

Table 11-1. Prisoner's Dilemma Payoff Table.

As the payoff matrix shows, Prisoner's Dilemma is a non-constant sum game with symmetric outcomes. The sum of the outcomes, years in jail, for the group is one to twenty years, depending upon Bob's and Sue's choices and how they combine. It is not a zero-sum game because the gain for one person is not necessarily mirrored by a commensurate loss by the other. The outcomes for each participant are parallel (their incentives and payoffs are the same), so the game is symmetric.

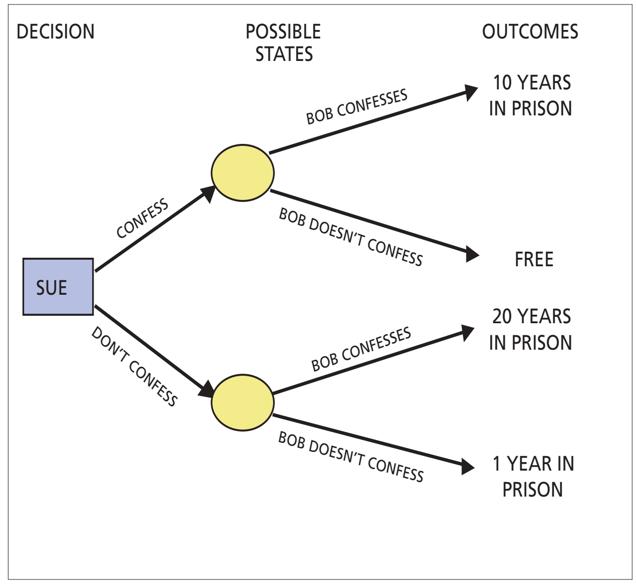


Figure 11-1. Prisoner's Dilemma Decision Tree.

Place yourself in Sue's situation. If she confesses and betrays her partner in crime, she will either go to jail for 10 years, or get probation. If she does not confess and refuses to incriminate Bob, she will definitely go to jail, either for 20 years or for one year. How should she decide whether to confess? Should she consider Bob's behavior since his decision affects her outcome? If Bob confesses, she should confess for the lesser penalty of 10 years in prison instead of 20. On the other hand, if Bob does not confess, Sue is still better off if she confesses because she will not go to prison at all. Her individually rational position then, is to confess. Sue's decision tree looks like figure 11-1.

Bob faces the same choices. His reasoning is identical to Sue's and he confesses, too. They both go to prison for ten years. Clearly, as a team, both would have been better off by not confessing and going to prison for a year each and being released young enough to enjoy their ill-gotten gain. However, since they are unable to communicate prior to or during the game, both rationally select the dominant personal strategy, the one that maximizes individual payoffs. How do the results of Prisoner's Dilemma help us understand complex negotiations? They illustrate clearly that individually rational behavior aimed at maximizing personal gain can result in worse group outcomes than *seemingly* irrational individual behavior.

Why, you might ask, would they both not see the possibility of helping each other by refusing to confess? Perhaps one or even both did. However, then they must ask themselves whether they can trust the other person to see the opportunity, too, *and* act upon it. Assuming there are no emotional bonds complicating the decision, without communication, the rational individual choice for either Bob or Sue is to defect from the group-beneficial strategy of silence and confess in order to avoid even a year in jail. Thus, even a prior arrangement to stick together may not hold after they are captured and must actually decide whether to confess. Does Prisoner's Dilemma imply that negotiations can never be undertaken in good faith? Not necessarily, but it drives home the importance of knowing, or trying to discern, what your opponents see as their best personal outcome, whether there is a more positive group outcome possible, and devising your own strategies accordingly. Prisoner's Dilemma also drives home the value of genuine mutual trustworthiness and clear communication.

THE TIT-FOR-TAT STRATEGY

By limiting the Prisoner's Dilemma to one move and prohibiting communication between the prisoners, each participant's rational choice is his or her dominant strategy. Bob and Sue opt to confess to optimize their personal outcome regardless of the other's move since that move is unknown. If we allowed the game to progress beyond one move, we may allow Bob and Sue to build the basis for cooperation if they use their moves to communicate their intentions. Professor Robert Axelrod of the University of Toronto iterated Prisoner's Dilemma to explore optimal strategies under this new condition with the help of game theorists in a round robin tournament.

They discovered that a strategy called Tit-for-Tat (literally "an equivalent given in return") is the most successful overall approach to Prisoner's Dilemma. Players begin executing this strategy by cooperating[2] on the opening move then mirroring the move of their opponent. Cooperative moves by the opponent then lead to mutual cooperation, while defection by the opponent results in both players defecting. The basis of a tit-for-tat strategy is the inherent communication of the move itself. If Sue were to employ this methodology with Bob, she signals him by responding in kind to his last move. She is telling him she will cooperate for mutual benefit if her cooperation is reciprocated. She is also telling Bob she is willing to retaliate if he will not cooperate.

The dilemma of all cooperative games is that at some point an individual player is tempted to defect in order to maximize individual gains. If the participants know exactly how many moves they will be allowed, they have a powerful incentive to cooperate until the last round.

TIT-FOR-TAT STRATEGY

- Cooperate on the first round to demonstrate a desire for mutual gains.
- On each successive round, copy what your opponent did on the previous round.
- On the last round, defect to maximize personal gain with no risk of retribution in the future.

They may then defect to maximize their individual outcome—without providing the opponent an opportunity to retaliate. The opponent, however, may anticipate this move and defect on the next to last move to preempt being so caught, and so on.

To prevent a series of preemptive defections from backing all the way up to the first round and keep participants in a cooperative mindset, there are several conditions we can create to achieve success with a Tit-for-Tat strategy:

- The relationship between negotiating parties must be open-ended; there can be no last turn. While the current issue may be resolved, there will be more issues involving these participants in the future.
- The future payoffs, those that will accrue from continuing cooperation, must be large enough to offset the immediate gratification of defection. This underscores the importance of understanding the value you and your counterpart place on possible outcomes.
- Do not get trapped in a competitive mindset by comparing your gains to your counterpart's to measure your success. In non zero-sum games, your payoff does not have to come at the expense of the other side.

In DoD, our negotiations cover a wide variety of topics but often involve the same people or organizations. We have a natural set of continuing relationships and therefore it is usually in our best interest to cooperate throughout negotiation to build long-term trust and confidence. Each of us has probably seen the chill that develops around an organization or individual that relies

2. In the case of Bob and Sue, the Cooperative strategy is “Don’t confess” because they are cooperating with each other (not the police); their Defect strategy is to confess.

on defection to maximize their individual gain and the lasting damage that a lack of trust causes when they try again to participate in negotiations. We should defect only when our cooperation results in a truly unacceptable outcome. Even then, our defection should be clearly based on our interests, not on maximizing a position at the expense of our opponents.

ASYMMETRIC VALUES

So far, we have illustrated the basic problem of achieving cooperation between parties who have a symmetric and mutually understood payoff matrix. Now we are going to introduce asymmetries in the value of outcomes in the payoff matrix. Sometimes, identical outcomes are more important to one organization than another. A $50 loss or $200 gain may not be as important to a wealthy individual as it is to a poor one.

This value or utility difference creates some interesting possibilities for bringing groups together. For example, when the Navy and the Coast Guard negotiate the defense features (paid for by the Navy) that will be incorporated in a new Coast Guard cutter, the features usually amount to a few million dollars for each ship. The Navy considers variations in these costs modest because the defense features of Coast Guard cutters are a very small portion of the Navy Shipbuilding and Conversion budget. The same changes in similar Navy ships would be barely noticeable and hardly discussed within Navy budget circles. To the Coast Guard, however, these changes and their spillover effects (like additional spare parts and crew manning) introduce costs to the Coast Guard that are a significant portion of their budget and will require sacrifices in other areas. If the Navy wants to add the features, it has to provide an incentive for the Coast Guard.

We can combine issues with asymmetric values to produce mutually acceptable agreements. Suppose we have a problem that has several viable solutions and we need another organization to agree on the alternative we will recommend; effectively, they have a veto. Imagine we are on the Navy Staff and we have invested in a data link architecture that requires specific protocols between using units. We know that the Joint Staff and the Office of the Secretary of Defense are going to review proposed standards for joint data protocols and we want them to adopt ours instead of two other candidates (X and Y) proposed by contractors. The Air Force is the other major stakeholder concerning this data format. Our counterparts on the Air Staff have not committed to a particular data protocol, but both contractors' formats are more compatible with their existing equipment and less costly (to the Air Force) than the Navy-preferred format.

We also know that Navy and the Air Force have different views on the air control procedures for interceptors. The Navy, using data links between fighter aircraft and controlling platforms, minimizes voice communication between units and relies on the data network to relay most of the tactical information to the interceptor aircrew. The Air Force relies less on data links for fighter control and often uses broadcast control procedures that are based on providing (voice) bearing and range calls from various kinds of reference points, one-way voice transmissions from controllers, and more local (autonomous) control of the intercept by the aircrew. If the Air Force accepted the Navy style of interception, they would require more and different equipment and expensive training. On the other hand, the Navy, although it prefers the use of data links, is comfortable using broadcast control.

A payoff table for the two staff organizations on the two issues looks like this:

	NAVY			USAF		
Alternative	**Cost**	**Benefit**	**Utility**	**Cost**	**Benefit**	**Utility**
Navy Data Format	Low	High	88	Medium	Medium	25
Contractor Data Format X	Medium	Medium	31	Low	Medium	32
Contractor Data Format Y	High	High	22	Low	Medium	38
Network (Data Link) Control	Low	High	75	High	Medium	20
Broadcast Control	Low	Medium	65	Low	High	89

Table 11-2. Navy and Air Force Preferences for Two Linked Issues.

By examining either issue in isolation, the Air Force and the Navy have clear individual preferences, based rationally on cost and benefit, which are contrary to the other services. There is no incentive for them to cooperate and support each other—until we consider both issues together. (Note the relatively small differences in utility value in table 11-2 to each service for its less important issue.) The asymmetric values of the alternatives can lead both organizations to agree on using the Navy data protocol and proposing Air Force (Broadcast Control) fighter interceptor procedures as joint doctrine.

VALUE CREATORS AND VALUE CLAIMERS

David Lax and James Sebenius[3] describe negotiators as members of one of two camps: value creators and value claimers. Value creators construct agreements that produce mutual gains for all parties while value claimers use a competitive or win-lose approach. Negotiators create value in a bargaining process when they make tradeoffs and introduce more issues to link to the problem. By expanding the basis for agreement, at little individual cost to each participant, value adders exploit asymmetry to garner consensus.

Value claimers tend to see this emphasis on joint gain as naive, weak-minded, and a failure to recognize that power is at least as important as legitimacy to achieve a negotiated outcome. To value claimers, if one side wins the other must lose. This observation is not a negative judgment on humanity; it is a pragmatic assessment of negotiation. Value claimers enter negotiations by overstating their position, disparaging others' concessions, and waiting out their opponents. They think any other position invites disaster and that concessions and tradeoffs equate to weakness.

Actual negotiation, formal or informal, includes value claimers and value creators. Furthermore, any single participant may assume either role under different circumstances. Few agreements are possible without some value-adding steps, and no agreement is brought to closure unless some value is claimed.

AVERAGING

A common negotiation strategy is to split the difference between the parties' positions. This technique may lead to agreement when the risks are too high to forgo a settlement or if the participants' positions are fairly close. It is also a convenient strategy if the value of the alternatives (costs and benefits) is low to each side. Cunning negotiators, particularly value claimers, may

3. See David A. Lax and James K. Sebenius's book *The Manager as Negotiator: Bargaining for Cooperation and Competetive Gain* (New York: The Free Press, 1986)

exploit the unwary with averaging by opening negotiations with extreme or unrealistic positions, hoping to compromise their way to a favorable outcome by splitting the difference or meeting in the middle between opening positions. If we prepare well for negotiations by understanding the interests of the other participants and the context of the problem, we will quickly identify and discard outrageous opening positions. This strategy can also backfire on its practitioner; by offering to meet in the middle, the negotiator has identified an acceptable agreement that effectively becomes their new position, far from the previous level.

SCORECARDS

Successful negotiation requires careful analytical groundwork. We accomplished much of this preparation during the Analysis Phase by using a rational approach to choose among the alternatives. The analysis and the models we constructed to make our decision are also valuable during reconciliation. Now, as we prepare for negotiation, we should make a scorecard for each participant, including ourselves. On it, we list the interests, positions, and how the different outcomes affect each participant. We determine which of the possible outcomes are acceptable to us, and their relative desirability, and then do the same for the other participants. The scorecard may be figurative or it may be a document. We accept that our evaluation of the other participants may be imperfect, knowing that we can update the scorecard during the negotiation. As we develop these scorecards, we ask questions like:

- What is our threshold for an agreement? What will cause us to walk away from these negotiations?
- What are their interests? Are the other parties more concerned about budgets, schedules, or effectiveness?
- Can we determine how they selected criteria and which they emphasized?
- Can we derive their interests by examining their initial position or the criteria they selected?
- Are they under pressure to reach an agreement?
- Where can they afford to give concessions?
- Where can we make concessions and will they have significant effect on others?
- How do the parts of this problem connect to each participant and are there other issues we can introduce to add value (or sweeten the pot)?
- How do these parts link to each other? What is the general form of the most likely branches and sequels of the negotiations?

Thus, we identify possible outcomes and targets instead of entering with vague intentions to do well. Using a system similar to evaluating criteria on the basis of their utility, as we discussed in Chapter 6, we identify the milestones that we will try to reach during reconciliation. Our scoring system may be based on dollars, billets, arbitrary points (utility), or any other methodology suitable to the situation.

The first milestone is the worst possible outcome, the result of negotiations failing completely and stakeholders making the worst possible decisions from our point of view. Next is the minimum acceptable outcome, the least desirable outcome we will agree to under any circumstances. Another threshold value for negotiations is our Best Alternative to a Negotiated Agreement, or BATNA. It is the best outcome that we can achieve in the absence of negotiations or if we cease negotiations without an agreement. As we prepare scorecards, we need to identify our BATNA carefully. Naturally, we should not consent to an agreement unless the value of its outcome exceeds our BATNA. Our BATNA may be above or below our minimum acceptable outcome; if it is below, then we are under great pressure to reach an agreement. Our scorecard will also contain the expected or most probable outcome and our optimal, target, or ideal outcome.

After we have assessed and listed our own range of outcomes, we turn to the range of outcomes for the other participants. For each, we "estimate" their minimum acceptable outcome, BATNA, expected outcome, and ideal outcome. We may be able to detect which other negotiators are under pressure to reach an agreement. Thus, we conclude who is negotiating from a position of strength and the basis of their strength, which may also help us detect their bluffs. There will be overlaps between our range of acceptable outcomes and the range of acceptable outcomes for the other participants. This overlap is called the zone of possible agreement; if there is no overlap at all, then there is no reason to negotiate.

Next, we craft an optimistic opening position, which is probably near our ideal outcome level because research shows that those who begin with positions near their goal are closer to achieving them when negotiations conclude. We prepare a list of arguments for and against our opening position and a list of arguments for and against the other side's expected opening position. When negotiating, naturally we present only our positive arguments while being ready to respond to counter-arguments. The other participants are likely to do the same.

BUILDING A SCORECARD: JOINT TARGETING

After the Gulf War, many in DoD recognized the need to better identify and attack mobile targets, e.g., Scud missile launchers, in order to reduce the time between their detection and attack. The Joint Staff hosted several meetings with the services in order to chart a way ahead to support the future procurement of time-critical targeting automated information systems that would fill this need. Because each service had for some time aggressively attempted to solve the problem individually (incurring much sunk cost in the process), the Joint Staff felt an urgency to establish defense-wide standards to ensure interoperability between them. A Joint Staff officer's scorecard before the negotiations might have looked like this:

Joint Time-Critical Targeting Standards Scorecard

Joint Staff J6 Directorate

Range of Possibilities from Negotiating

- **Worst Possible Outcome**

  Each service develops duplicative and non-interoperable software.
- **Minimum Acceptable Outcome**

  Services agree to liaison with counterpart program offices to ensure interoperability.
- **Most Probable Outcome**
  - Services agree to do more in-depth exploratory analysis before committing to future development.
  - Services do a competition selection of software to create a single joint standard.
  - Services sign a Memorandum of Agreement to share funding responsibilities.
- **Best Possible Outcome**
  - Identify a service as the Executive Agent for all DoD time-critical targeting software.
  - The Executive Agent subordinates themselves to a Joint Staff-led requirements and configuration control process to ensure each service's needs are met.
  - Each service agrees to ignore sunk costs and only use the new joint standard software for future programs and retrofits.
  - The Executive Agent funds all future software developmental costs.

Best Alternative to a Negotiated Agreement

- Status Quo: Each service continues to develop its own duplicative software.

Service Interests (The Army, Navy, and Air Force all had similar interests.)
- Protect existing programs (sunk costs).
- Maintain control of their budget authority.
- Obtain funding support from other services.

Service Positions
- Offer to become the DoD Executive Agent – including funding and standards oversight – for all DoD time-critical targeting software.
- Existing programs are too far advanced to be abandoned.

JS J6 Opening Position (Same as Best Possible Outcome.)
- Identify a single service as the Executive Agent for all DoD time-critical targeting software.
- Consolidate DoD funding under the Executive Agent for all future software developmental costs.
- Push the services to accept a Joint Staff-led requirements and configuration control process to ensure each service's needs are met.
- Push the services to agree to use a single standard, to be determined, for future time-critical target information processing programs and retrofits.

RECONCILIATION TACTICS, TECHNIQUES, AND PROCEDURES

Throughout negotiations, we keep track of each party's current proposal and how their proposals have changed over time. We are constantly seeking opportunities to move closer to an agreement. As we take part in negotiations, we use these rules of thumb to achieve an optimal outcome:

- *Find weaknesses.* Determine what is pressuring the other participants for an agreement, where their interests are vulnerable, and then exploit them or openly help to protect them to gain leverage.

- *Do not present the other side's position.*
- *Delay*. Ponder every little thing. Be prepared to use all the time available, even when agreement is close early in the process, because negotiators tend to concede more at the end of negotiations when there is a risk of losing the agreement. Do not settle too quickly unless we have our goal. Fast deals may include dangerous oversights.
- *Proceed incrementally*. Control the process and make many little moves rather than a few sweeping ones so we can carefully track our concessions and gains.
- *Collect intelligence.* Test the assumptions we made to build our scorecards. Ask for information from the other parties such as worksheets and studies. Reveal as little as possible about our own pressures.
- *Reduce risk*. Spread risk among the participants so that it is not dangerous for any one participant to buy into the agreement.
- *Consider legitimacy.* Data can come from many sources, but the way it is packaged can give the impression it is more legitimate, e.g., bound, glossy books, instead of hand-scribbled lists and sketches.
- *Be careful directly contradicting another party.* Challenges invoke ego responses that may shut down negotiations.

CONCESSIONS

Handling concessions is an inevitable part of negotiation and a delicate skill that we need to develop. There are several general principles we should consider as we modify our negotiating position. If another participant makes a large concession early, it is an indication he or she probably can still give a lot more. Our concessions should be small, giving the impression we are already close to our BATNA. When we decline an offer, any offer—even if it meets our requirements—the other participants are likely to wrongly assess our position and concede more to reach an agreement.

Concessions should appear painful. To preempt further requests, we should make our minor concessions appear to be very important before we present them. We should try to make the representatives from other organizations appear successful. Giving in slowly makes the other negotiators look like they did a good job. Our concessions should be infrequent because we make them only to get the reconciliation moving again. We should make participants compete with their offers, i.e., by telling each of them (individually) they must do better to satisfy us (the "krunch" tactic).

For every concession we make, we should try to maximize our return and minimize our cost, e.g., "we will provide the instructors if you fund their travel." Naturally, we make concessions in areas that have less value to us and more value to the other participants. We may refer to a "third party restrictive force" outside the negotiation that limits our ability to concede more, e.g., " This appears feasible, but my superior has to approve it."

We should be especially careful as the deadline approaches. Big movements happen as participants weigh the risk of no deal against their BATNAs and the time and effort they have already invested in the negotiation. When an agreement is wrapping up, search for small additional concessions, e.g., free documentation, consulting, shipping, or warranty repairs to reduce total ownership cost. We may expect the other participants to employ these same tactics. For example, if we successfully nibble at the end of an agreement, we may find each new item listed in a detailed invoice marked no charge to ensure we "know" the value of each concession.

ATTACKING EXPECTATIONS

During the negotiation, we want to convince the other participants that they have set unreasonable milestones and that the final agreement is likely to be much further toward their minimum acceptable outcome (and our ideal outcome level). If they want an agreement with us, they will have to make bigger concessions than they had planned. Any number of negotiating tactics may affect the other participants' expectations about possible agreements, even if the effect is only psychological. For example, "flinches" are deliberate body language or verbal reactions we use to attack other participants' expectations, such as, "That's more than double what we budgeted!"

Patterns of success and failure influence negotiators' expectations and those expectations affect the final shape of an agreement. Negotiators raise their goals after achieving a success (they expect they will get an even better agreement the next time), and they lower expectations after failures, especially if they attribute the success or failure to their own performance. We need to be sensitive to the mindset of our team members to ensure that our positions are tied to our interests and not to our moods.

VALUE PERCEPTIONS

Some negotiation techniques do not really change the value of the agreement; they change the perception of the value of money or give the impression that a concession is being offered. This tactic is called "funny money." Its objective is to reduce the other participant's anxiety about cost or to show how much more the participants are getting for their money. Contractors in reconciliation situations with DoD negotiators may use "funny money" to try to hide costs, as listed below:

- List an average cost instead of total cost, e.g., unit price compared to the total cost of the purchase.
- Transfer duties from the civilian labor force to the uniformed staff and exclude the cost of additional military labor because servicemen are not paid overtime.
- List the procurement price instead of life cycle costs.
- List incremental payments, e.g., outlining monthly payments instead of the total purchase cost.
- Use units of measure other than currency, e.g., man-hours.
- Display budget reductions by percentages instead of actual dollars.
- Use non-standard measures, e.g., "truckloads of parts" instead of a dollar amount.

DEADLOCKS

Before we conclude our negotiations without an agreement, there are several steps we can take to spur movement in positions. We can try to find common ground on a personal level with other negotiators to build trust and confidence. We can search again for new information or another issue to add value. We can change negotiators, either to adjust personalities or to introduce more senior personnel. The senior negotiators may also separate themselves and take "a walk in the woods" for a private, protracted discussion to get things moving again.[4]

---

4. Strobe Talbott, *Deadly Gambits*, rev. ed. (New York: Vintage Books-Random House, 1985), Chapter 6. During the series of strategic and intermediate-range nuclear arms control negotiations of the 1970's and 1980's, the leaders of the U.S. and Soviet delegations, when they felt the talks were stalemated, would meet apart from their advisors and staffs at locations where both were safe from eavesdropping, electronic or otherwise. Many of these conversations led to dramatic shifts in positions and allowed negotiations to continue that were on the brink of failure.

Public Dispute Resolution

Problems do not always arrive in our in-basket as neatly staffed recommendations supported by careful analysis. Sometimes we learn about them first in the newspaper or hear about them on the evening news. Our course's Executive Decision-Making Framework is a good road map for crisis action as well as for deliberate problem solving. With problems in the public domain, we begin the Definition Phase knowing that we have a reconciliation-intensive situation. In this section, we present some additional considerations for decisions involving the public, angry or otherwise, and some examples of attempts at reconciliation.

The Department of Defense operates largely beyond the sight of the American public. In the past, DoD occasionally abused its need for operational and information security as a screen for inappropriate activities. Many of these breaches of trust, e.g., nerve gas testing, medical research on prisoners, and exposure of servicemen to atomic radiation, came to light many years after their occurrence and only after prodding by researchers using the Freedom of Information Act. Each exposé reduced DoD's credibility and made the public distrustful and ever more ready to believe new accusations.

When something visibly goes wrong involving the Defense Department, mutual (if benign) ignorance and our poor public affairs management often stymie cooperative relations with the public. In some cases, people are angry because they perceive that DoD neglected them or caused them harm. They demand restitution or compensation and they may threaten to file lawsuits. Senior defense executives confront this very uncomfortable situation more often as information, accurate and otherwise, is disseminated faster and faster with less scrutiny. In 1997, based on misinformation on the Internet, nationally syndicated columnist Pierre Salinger reported that a Navy ship shot down a commercial airliner, TWA flight 800, over Long Island Sound with surface-to-air missiles. Less dramatic public confrontations about environmental quality, noise pollution, waste disposal, equal opportunity, sexual harassment, and many other issues are becoming commonplace for DoD installation commanders and personnel managers.

THE DEFENSIVE APPROACH

Few DoD leaders intentionally take actions that are harmful to the public. Therefore, when many decision-makers are accused of such activities their reaction is surprise, shock, confusion, or even outrage. In some cases, DoD leaders have therefore reacted along the lines of what we call the Defensive Approach. The defensive approach is essentially negative and adversarial in tone. It is better suited to damage control than confidence building.

The Defensive Approach may be necessary when there is no common ground or DoD is subject to liability claims. But even in these situations, there is never a good reason to treat the public with contempt, or to lie. In the long run (and often in the short run as well), such behavior is bad for DoD, for your organization, and for our country. Therefore, we recommend that you distinguish between two versions of the defensive approach.

We call one variant the "hardline" defensive approach. It can involve some or all of the following actions in dealing with an angry public:

- Vigorously *deny that anyone is being harmed.*
- If someone has been harmed, *deny responsibility*. Convince the public that we are not responsible and our actions (or inaction) are not at fault.
- *Minimize the harm.* Hire experts to support our point of view and publicize their reports.

- *Stonewall*. Deny interviews and make bland, reassuring comments.
- Produce a *scapegoat*. Place the blame on an individual—preferably one who transferred recently—who acted against the wishes of leadership and the culture of the organization.

Figure 11-2 contrasts the hardline defensive approach with the variant that we recommend.

"HARDLINE" APPROACH	RECOMMENDED APPROACH
Deny that anything's seriously wrong	Do not concede that anything's seriously wrong
Deny responsibility	Do not admit liability
Minimize the harm	Downplay the issue if you can do so honestly
Stonewall	Wait for public attention to shift
Scapegoat	Accept individual responsibility sacrificially

Figure 11-2. Two Variants of the "Defensive Approach"

The approach that this course recommends avoids the danger of making assertions that might not be true. Instead of denying that's anything is seriously wrong, we recommend that you explicitly reserve judgment until you have seen more evidence. Instead of saying that DoD or your organization is not responsible, we recommend that you take care not to admit liability. Instead of minimizing the harm, we recommend that you downplay the issue (e.g., by focusing on other issues, or otherwise trying to shift the terms of the debate). Instead of stonewalling (e.g. by making reassuring comments that may or may not prove well-founded), we recommend that you say as little as possible and wait for public attention to shift elsewhere. Instead of finding a scapegoat, we recommend that you do nothing. (A related but honorable option is for an official to accept overall responsibility on behalf of the organization—and perhaps even resign as a kind of sacrifice—even if he or she is not directly or individually at fault.)

CASE STUDY: TAILHOOK—THE "HARDLINE" DEFENSIVE APPROACH

For the most part, the Navy's handling of the 1991 Tailhook affair is illustrative of illustrates the "hardline" defensive response to an angry public. When the news broke nationally of the alleged sexual harassment by and misconduct of many officers at the annual convention of naval tactical aviators in Las Vegas, how did senior Navy leaders, including some present at Tailhook, downplayed the reports respond?. They essentially denied that anything seriously wrong had happened and said that Tailhook 1991 was no different than past years (deny harm).

As the story persisted and accounts became more graphic, Navy leadership said that the incidents were confined to a specific locale, the third and sixth floors of the Hilton, not the whole conference. They claimed only a small number of the conventioneers participated (deny responsibility) and that all the incidents were consensual, or at worst, were misunderstandings (minimize the harm).

As the press accounts persisted and a female naval aviator reported she and others had been assaulted in a crowded hallway, the Navy's senior leaders dug in and announced a sweeping investigation was underway and therefore they could not comment on an ongoing investigation (stonewalling). The aviators under investigation, primarily Commanders, Lieutenant Colonels, and more junior officers, sensed that the Naval Criminal Investigative Service was looking for scapegoats and they generally refused to cooperate with the interviewers. Lawyers for many of them actively provided leaks to the media and held press conferences alleging a cover-up. They purported their clients were being smeared by the women who had taken their accusations public or by the other aviators being investigated. In the end, none of the Tailhook investigations indicted any officers for courts-martial.

Still, the publicity did not stop; the media accused the Navy of whitewashing and Congress demanded accountability. The Secretary of the Navy resigned (more scapegoating), but that was not enough. For the next five years, Congress reviewed all officer promotion lists for possible Tailhook participants. They went so far as to require sworn statements from each officer selected, regardless of seniority or job description, that he and none of his subordinates had participated in Tailhook 1991, slowing promotions throughout the Department of the Navy. If an officer could not so swear, his promotion was placed on hold for further detailed review with in light of Tailhook investigation files.

Many Naval officers were disappointed in their senior leadership for not facing facts openly, for failing to acknowledge responsibility (it was common knowledge within the Fleet behavior like this occurred at many previous Tailhooks), for not conducting an impartial investigation, and for preventing Draconian reactive measures by Congress. The spill-over continued; critics say Congress passed legislation without an extensive debate that expanded the role of women in combat as a form of compensation for Tailhook.

The Tailhook scandal was a public relations and policy nightmare for the Navy that could have been handled differently using the concepts for mutual gains to induce cooperation during the resolution of this dispute. When our personnel or we ourselves behave badly, there has to be an accounting to the public. We should strive to make this accounting in a way that punishes malfeasance, minimizes damage to the organization, and restores public confidence.

THE MUTUAL GAINS APPROACH[5]

A mutual gains approach will not make negative events or bad publicity disappear, but it can minimize their impact and it sometimes results in long-term benefits for the organization. Instead of confrontation, this approach seeks to structure negotiation with the public as a collaborative problem-solving effort with participants from all sides so that each has a better outcome. Our organization can apply the mutual gains approach by taking the following actions:

- Take the *initiative*. Do not wait to be put on the defensive.
- *Acknowledge* the concerns of the other side—they, too, believe they are right.
- Encourage *joint fact-finding*. Contradictory expert opinions cause the public to dismiss facts because they believe an "expert" can be found for any point of view.
- Insist upon *objective criteria.*
- Build *coalitions* to isolate the uncooperative.
- Seek *consensus*. Give others a reason to do what we want them to do rather than convince them that they are wrong. Analyze their interests and find areas for mutual gains that provide an incentive for cooperation.
- Offer *contingent commitments.* These may help to alleviate public concern about effects, e.g., "If we really are harming the environment, then we will take corrective action in concert with the community." We must be very cautious not to overstep our authority

5. The term "Mutual Gains Approach" and many of the negotiating ideas we present in this section are derived from work done by the MIT-Harvard Program on Negotiation. See, for example, the article "Dealing with an Angry Public" by Lawrence Susskind and Ira Alterman in the *New Jersey Bell Journal*, Fall/Winter 1991, p. 35.

when we make commitments; few of us can commit the U.S. government to a course of action.

- Maintain *mutual trust.* We should always act in an honorable fashion; our word is our bond and we do not make promises we cannot or will not keep. Once trust is lost, it is very difficult to regain.
- Accept *responsibility* and admit mistakes - but be careful about liability.
- Emphasize *outcomes*. Do not lose sight of the long-term objectives.
- Focus on long-term, *continuing relationships*. Short-term victories obtained at the expense of long-term interests are seldom worthwhile.

When outside activities reveal problems involving DoD, solving them is often in the mutual interest of the public and DoD. We can resolve these disputes more easily when we have principled, decent, and honorable dealings with the media and in our negotiations with civic leaders. An open approach to conflict resolution by Defense Department leaders will help maintain or restore the public's confidence in us. We can lay some of the ground-work for a successful mutual gains approach to a public dispute by fostering an open, cooperative relationship with our local communities. An aggressive public affairs plan is a useful tool in this regard.

Remember always, mutual gains means that our organization gains, too. It is not a recipe for negotiating concessions or a strategy based on weakness. Rather, a rising tide lifts all ships.

CASE STUDY: THE AIR FORCE ACADEMY[6]— THE MUTUAL GAINS APPROACH

Lieutenant General Hosmer, U.S. Air Force, was the Superintendent of the Air Force Academy in February 1993 when female cadets made a series of allegations of sexual assault. The sequence of events began when a woman freshman reported to her chain of command that several male cadets had sexually assaulted her outside the gymnasium. Soon afterward, a dozen more women came forward and reported incidents as serious as rape.

Within two weeks of these reports, General Hosmer called a meeting with all women cadets and promised complete confidentiality and no retribution in exchange for ground truth from the assembled women. The meeting lasted for four hours and he received a full appreciation of their accumulated grievances. General Hosmer next met with all the male cadets in a similar forum. While half of the women knew of sexual harassment incidents and assaults, ninety percent of the males were unaware of these problems at the Air Force Academy.

Armed with a better understanding of the problem, General Hosmer took a multi-pronged approach. He brought in investigators to handle the formally reported incidents, set up counseling for the victims, established a 24-hour hotline for reporting incidents, and increased sexual harassment awareness training. Two perpetrators were ultimately jailed, three resigned from the academy and three others were disciplined but allowed to remain. This distribution of outcomes indicated to the women that sexual harassment would be punished and to the men that each case would be dealt with individually, i.e., the punishments befit the crimes.

General Hosmer used a mutual gains approach to solve his problem effectively. He acknowledged the concerns of those involved, engaged in joint fact finding, clearly communicated his re-

6. Based on the account authored by the *New York Times,* "Air Force Academy Acts to Curb Sexual Harassment," reproduced in the *Providence Journal*, 2 May 1992, p. A16.

sponsibility and intent, maintained trust, intelligently accepted responsibility, and selectively shared power. All the while, he focused on the long-term relationships essential to his command. There are no protracted lawsuits as a result of these incidents and the Air Force Academy became a more effective institution in the long run. The command has moved on.

CURRENT APPROACHES

In any public dispute, there is a natural tension between each party's desire to cooperate to solve a problem and create value versus their desire to compete, protect themselves, and claim value. How we react as senior leaders in DoD depends on the situation and, as in any negotiation, we may switch strategies and combine tactics. Actual liability or operational security may force us to be unyielding, obstinate, and defensive at times. We know that behaving in that manner crystallizes the opposition and bonds them more closely against us. As rational trustworthy public servants, we should use mutual gains approaches to reach out to moderate elements of a dispute when we have common interests. We hope then they will take a less strident stand, bring along the media, and undermine the more radical fringe as we find long-term solutions to the underlying causes of public disputes.

CASE STUDY: THE RECONCILIATION PHASE
USMC MEDIUM-LIFT REQUIREMENTS: THE V-22 OSPREY AND HELICOPTERS

Many groups had positions on the purchase of V-22 Ospreys before the Institute for Defense Analyses (IDA) Study was completed. Secretary of Defense Cheney had already decided that DoD should not buy the V-22 because it was too expensive and should instead purchase less expensive helicopters to solve the Marines' medium-lift problem. Many observers and participants expected that IDA, who conducts many studies for the Office of the Secretary of Defense and other defense agencies, would automatically support Secretary Cheney's position. Instead, IDA concluded, as did the six earlier cost-effectiveness studies, that the V-22 was the best alternative.

The Office of the Secretary of Defense did not accept IDA's recommendation and Secretary Cheney and Dr. David Chu, which was Assistant Secretary of Defense for Program Analysis and Evaluation, remained opposed to the procurement of V-22s. While both acknowledged the V-22 outperformed helicopters, they felt it was too costly in the context of the Department of the Navy's overall program requirements in the near term. DoD needed money in the next set of budgets for other programs in addition to Marine medium lift. Meanwhile, the Marines, with the backing of Congressmen whose districts included the V-22 manufacturing plants, welcomed the IDA study and its recommendation.

The reconciliation process came to a head in testimony before Congress—which had directed DoD to commission the study in the first place. Dr. L. Dean Simmons, IDA's director of the study, presented his results and Dr. Chu testified next. Dr. Chu pointedly criticized the study's assumptions and recommendations but acknowledged, despite his concerns about maintenance and sortie rates, that the V-22's performance was clearly superior to any of the helicopter options.

Dr. Chu said, "The bottom line here, sir, with great reluctance by the Department [of Defense], is, we cannot afford to spend the kind of money that starting this [V-22] production line and buying these aircraft in reasonable numbers would require….

"What this would compel the Department to confront are a series of very painful tradeoffs to find the several billion dollars necessary to sustain that buy, not only in the period 1991-1997, but in the years beyond. [We question] whether we have enough money to buy both the ships that the Marines need and the aircraft, if we go for an elegant aircraft solution [the V-22]….

"[We must] avoid letting 'better' be the enemy of 'good enough.'"

Essentially, Dr. Chu, who had to apportion the entire Defense Department budget among competing programs, felt the V-22 imperiled other, more necessary programs.

Senator Dan Inouye (D-HI), chairman of a Defense Appropriations Subcommittee, commented he had been a member of the subcommittee for about twenty years and this was the first time he could remember that the Office of the Secretary of Defense had attacked an IDA study. After more robust questioning, particularly by Senator Specter, Dr. Chu and the Senators came to agree the fundamental issue was near-term cost. (The record of this testimony is in Appendix 3; it provides great insight into the important role of analysis preparing our senior leaders for congressional testimony.)

Congress provided the DoD participants with a mutual gains solution by adding value. They provided new money to fund the V-22, increasing the amount of resources available to DoD. But the story still was not over... Secretary Cheney refused to obligate the money Congress provided - he did not want to start a program whose whole life cycle cost he thought too expensive. Congress was taking him to court to force him to spend the funds they authorized and appropriated for the V-22 when the 1992 presidential election made the point moot. The production of the V-22 seemed assured….

Summary

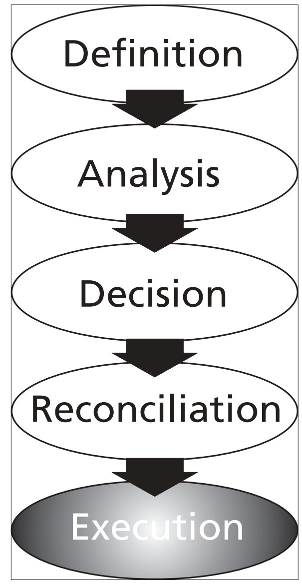
All of our defense-related decisions must be reconciled among competing interests before a decision can be executed. We believe that the differences among preferences are best reconciled through negotiations and that mutual gains is usually the preferred approach for reaching an agreement. For those instances where there is no common ground, we recognize a zero-sum (defensive) approach may be necessary, and we may have to resort to adjudicators higher in the chain of command, traditional power brokers, and coalition building.

In this chapter, we have applied our Executive Decision-Making Framework to advance our organization's interests by negotiating for the acceptance of our preferred alternative. We examined how participants behave in negotiations, why they do so, and discussed two basic approaches to negotiations: defensive and mutual gains. Just as a good commander prepares for a campaign, we described some important preparations for negotiations like using analysis to build scorecards. While emphasizing mutual gains as the most beneficial approach to reconciliation, we provided several guideposts and principles for conducting negotiations, dealing with concessions, and breaking deadlocks. Finally, we looked at how we can use negotiation techniques to resolve problems with the public.

We believe that principled negotiation is an important skill and ethical responsibility of senior DoD leaders. Executives who fail to prepare for negotiations, approach them casually, or appeal to some other authority to resolve differences are inevitably bested by those who come prepared, understand the issue, know their interests, support their positions with analysis, and adjust their positions to find an acceptable path toward their objective.

CHAPTER 12

EXECUTION



Be always sure you are right—then go ahead.
—Davy Crockett, Autobiography, 1834

NOW THAT THE DECISION IS MADE AND RECONCILED, we have an alternative that we are ready to implement. The Execution Phase is where the program or policy becomes operational on time, within budget, and fields a system, changes a process, or achieves a policy effect. Most decisions that unravel do not fail because of hostility or opposition to implementing an alternative, rather because the organization plans inadequately, manages or oversees poorly, or fails to enforce requirements and standards.

After a course of action is agreed upon and funded, good decision makers direct their attention to meeting objectives, specifications, schedules, and budgets. The process may seem overwhelming, but there are tools available to help manage complex, interrelated tasks. In this chapter we extend the Executive Decision-Making Framework to encompass implementing a force planning alternative and verifying that what we plan is actually accomplished.

Implementation

The Execution Phase begins as we plan how we will commit the money, material, time, and personnel resources necessary to field a system or execute a policy. We consider three essential aspects. The first is to describe, plan, and schedule tasks. The second is to identify or create an organization and make it responsible for executing the alternative according to the plan. Third, we install a control process to ensure the implementation is done according to the plan. Many graphic techniques and computerized aids may assist us planning, organizing, monitoring, and controlling the process of implementation. These planning methodologies range from informal outlines to highly structured, technical computer programs.

A PLAN FOR IMPLEMENTATION

The implementation plan is the road map for carrying out the decision. Although the plan changes throughout the life cycle of the system or policy, our initial planning efforts remain important. Many facets of implementation planning occur well before the decision is made, e.g., cost, schedule, performance, and risk are usually part of our formal analysis. Performance goals, cost estimates, and the time projected to field and test a system probably affected our preference among options. Now they can serve as an outline for execution planning.

The criteria we used to assess alternatives are often appropriate controls for execution. The systems approach to analysis gives us a useful set of ideas about what must be done, when to do it, and the risks involved as we implement an alternative.

AN ORGANIZATION FOR IMPLEMENTATION

Regardless of which planning system we select, we must identify the people who will carry out the tasks. Who will do the contracting? Who will check on fabrication? Who will oversee and integrate the training? Who will monitor the system's progress? The answers to these questions characterize the responsible implementing organization. An organization may already exist that is capable of executing this project. For simple programs and small projects, ad hoc organizations are adequate. Large projects may require a new formal program office with hundreds of people. The decision maker must approve the organization, define its responsibility, and delegate the authority to execute the alternative.

CONTROLS FOR IMPLEMENTATION

After creating a plan and identifying the force planning organization that will execute it, the next step is to install a management control system to monitor the implementation progress. This control system monitors three critical factors: actual timing of scheduled events, levels of performance, and cost. The control system regularly compares the status of these facets against the program or policy objectives. When deviations from the plan occur, the executing organization fixes the problem. During the life of the program, trade-offs frequently take place between cost, schedule, and performance objectives. A well-designed management information system is a key part of this control system for monitoring trade-offs during implementation.

In essence, the Defense Acquisition System (see Chapter 4 in our *Resource Allocation: The Formal Process*) is an implementation, control, and monitoring system, one that is closely related to the world of analysis. After a need with a material solution emerges from the Requirements Generation System, it enters concept development in the acquisition process. Proponents compete different alternatives against each other as we described earlier in the Definition and Analysis Phases. Once a project matures and the concept narrows, the program managers craft the documents of the formal process, e.g., the Capstone Requirements Document and the Operational Requirements Document. They codify the Key Performance Parameters, Thresholds, and Objectives that will be quite similar, if not identical, to the criteria that enabled the decision itself.

The Acquisition Program Baseline and Acquisition Strategy are key management tools that program managers create to aid implementation. They are iterative (reviewed at each of the milestones) and self-regulating processes that constantly focus and re-focus the implementation of the decision on the mission need. The acquisition process is punctuated by these milestone reviews; after each, a program passes into its next formal procurement phase. The decision makers conducting these reviews may consider many issues beyond the program under review and its supporting analysis.

AIDS TO IMPLEMENTATION

The decision maker will find many tools and techniques in management textbooks that will help him or her implement a major decision. Some are more appropriate than others for a project. Most of these aids are available in computer programs but they may also be used efficiently for less structured manual applications. We include several representative methods here to familiarize you with their attributes; you will probably observe that many have applications beyond decision making and project management. These aids are often the basis for presentations and reports to senior defense executives, therefore it is important we understand what kind of information they do and should contain.

Flowcharts

A flowchart is a schematic drawing that shows the steps of a process and how they interact. We place the steps inside geometric shapes that show their function and we connect with arrows to indicate their sequence. As shown in figure 12-1, ovals mark the beginning or end of a process, rectangles describe activities, and diamonds show decision or inspection points. The numbers on the left are days into the process. Flowcharts may be used to plan an activity or to compare actual processes with ideal processes. They can also point out measurement points and duplicative steps. They are not good for showing steps that can be performed concurrently.

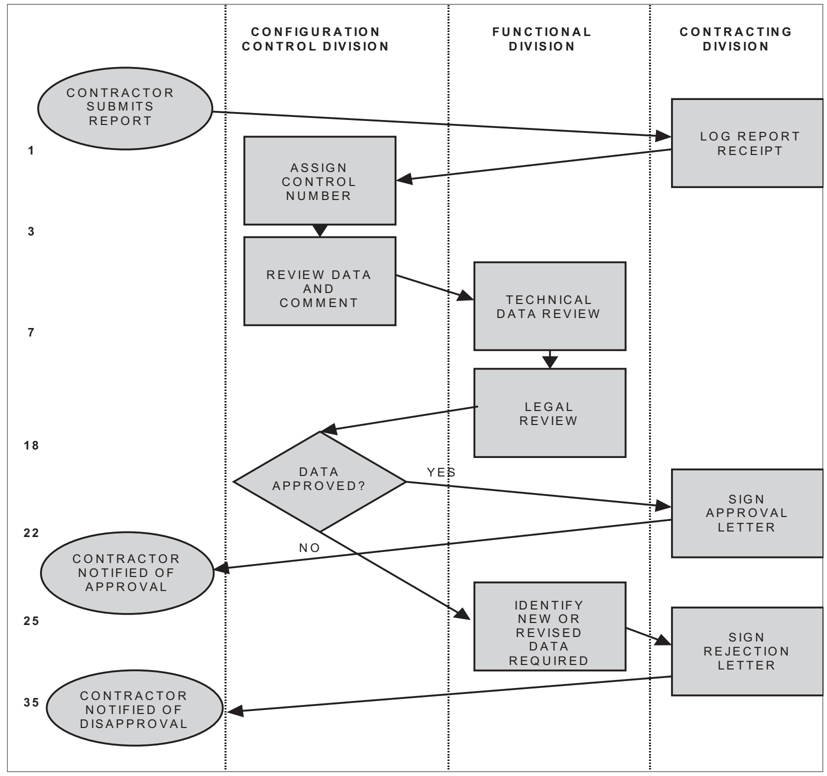


Figure 12-1. Correspondence Routing Flowchart.

Displays for Large Collections of Data

After we begin measuring, we may find that we have an overwhelming amount of data to sort. There are several ways to distill large amounts of data. A Pareto Chart is a bar graph that presents data in its descending order of occurrences. This type of chart focuses attention on a few significant events by separating them from a much larger batch of data with many insignificant ones. It is useful for displaying the effect of changing a policy. Careful selection of what to measure, as with any criterion, is very important because many statistics are misleading. For example, graphing the number of injuries in each department of a command may lead us to focus our safety effort on the department with the most injuries. It is possible, however, that department has the most injuries simply because it is the largest department while another department has many more injuries per person or man-hour. It is also possible that the injuries in the largest department are relatively minor while another department has injuries that result in much more lost time. It may be more appropriate to graph lost time due to injuries per capita for each department, as shown in figure 12-2, if we want to focus our safety attention on the area with the most impact on the organization or verify we have an effective policy.

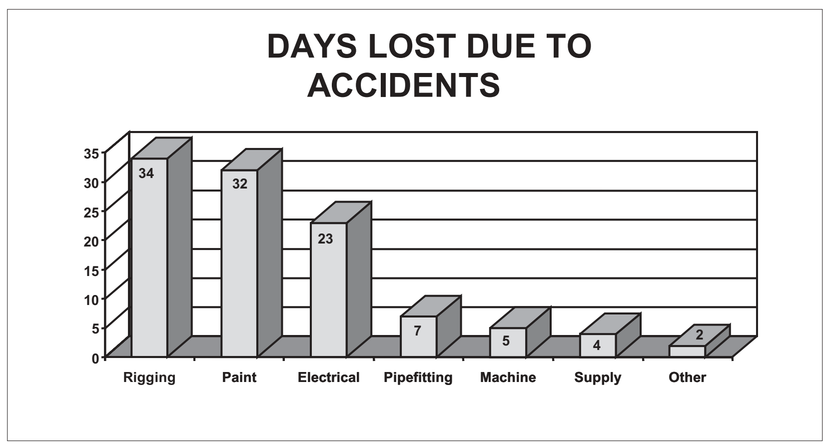


Figure 12-2. Serious Accident Pareto Chart.

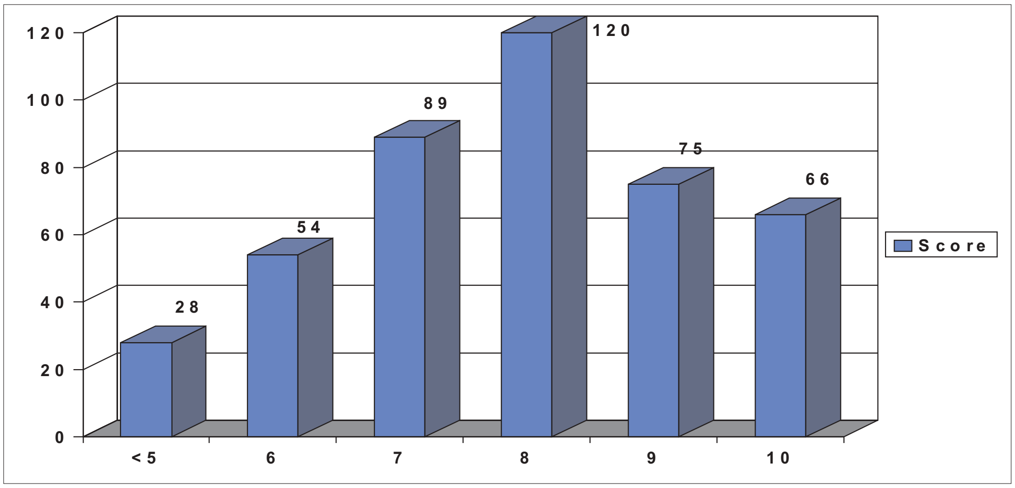


Figure 12-3. Marksmanship Histogram.

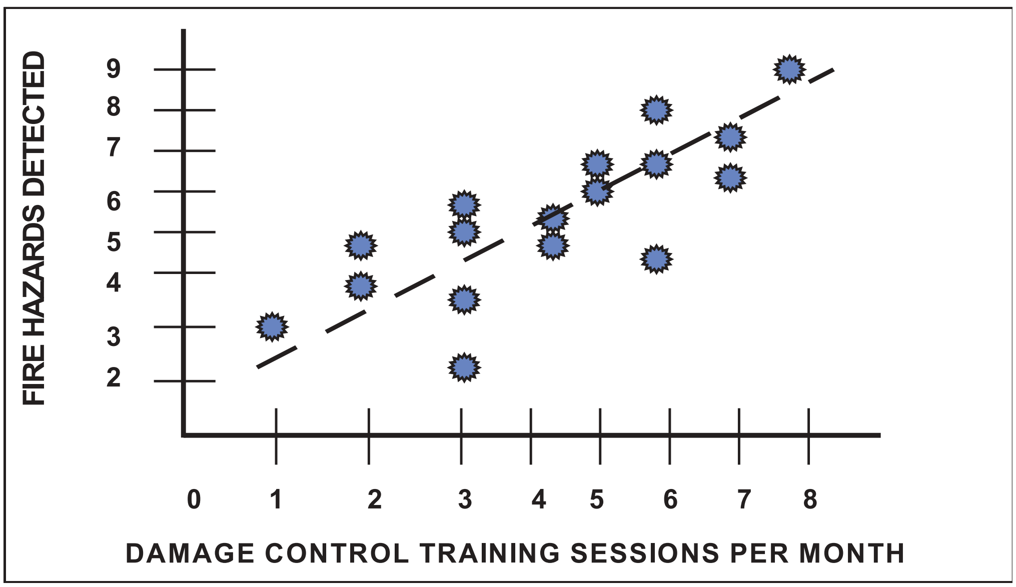


Figure 12-4. Hazards vs. Training Frequency Scatter Diagram.

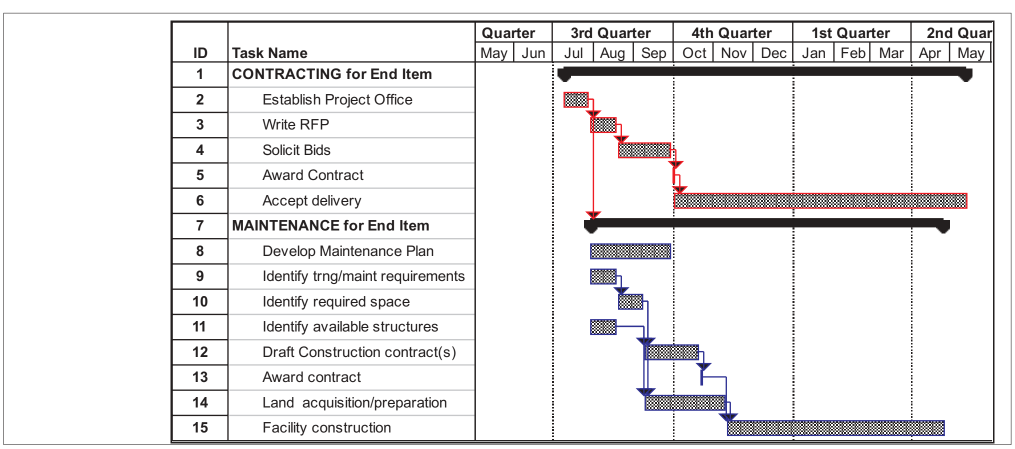


Figure 12-5. Procurement Program Gantt Chart.

Histograms are like Pareto Charts, but they are arranged to show the frequency distribution of data over a range of values. We place the largest concentration of data in the center of the range and values with less frequent occurrences in bins to either side. Histograms readily show the amount of variation in the data set and its distribution. Figure 12-3 is a histogram of marksmanship results for riflemen who score between 0 and 10 on their tar-gets.

Scatter Diagrams contain a raw plot of two variables related by cause and effect, for example fire hazards detected versus damage control training sessions. We seek to know if a change in one variable has an effect on the other. We plot the raw data on the chart with the independent variable on the x-axis and the dependent variable on the y-axis. After all the data is plotted, we construct a line of best fit through the data field to see if there is a trend that connects the two variables as shown in figure 12-4.

Gantt Charts

These charts list tasks vertically on the left side and their schedules graphically (horizontally) along the right side as shown in figure 12-5. The timelines for each task display its start and completion dates. We indicate slack time with dotted lines. Tasks may be broken down into sub-tasks and, for large projects, managers may create a series of charts that show increasing levels of task detail.

Gantt charts are very helpful tools for managing the Program Evaluation and Review Technique—PERT. They easily represent which tasks depend on the earlier completion of other tasks in a daisy chain effect. By building such a chart for the entire implementation, we can identify the Critical Path from the start to finish of the project. The critical path is the sequence of interdependent tasks that have no slack time between them; a delay in any of these tasks moves the completion to the right (further in the future). Gantt charts also make it easy to identify tasks that may be performed simultaneously or non-sequentially.

Critical Path Management has become its own discipline. If more resources become available, managers may apply them to the critical path to finish the project sooner, which will mean savings in fixed cost and possible performance incentive awards. If a task along the critical path is slipping, managers know this is an immediate cause for concern. Industrial activities, such as shipyards, use Gantt charts extensively for construction and overhaul estimates, planning, and management. There are computer programs that generate Gantt charts quickly and we can vary their displays and their level of detail easily.

We may use Gantt charts to graph our resources. Instead of listing tasks to be accomplished, we list our resources (personnel, equipment, contractors, etc.) on the left side and on the right indicate when they are involved with various activities. This prevents scheduling a unique resource simultaneously for two different projects. Computerized programs can link the resources to the project. If the project is rescheduled or delayed, all of the resources assigned to that project automatically reschedule, and the computer flags conflicts for resources. Most fleets and type commanders use a form of Gantt chart to schedule ships and squadrons. If several ships are involved in an exercise that is delayed by a week, the computer indicates which ships remain available to participate and which ships have conflicts with another commitment.

Activity Network Diagrams

An activity network diagram is a variant of a flowchart that incorporates time, the critical path, and all the tasks required to complete the project. It uses a series of circles (events), arrows (processes), and numbers (duration) to show the sequence and relation of activities in a project. The critical path is the sequence of events that determines how long the project will take; each delay in a critical path activity lengthens the overall project by a like amount. For example, in figure 12-6, the critical path follows the bold arrows from Event 1 through Events 2 and 5 and ends at to Event 6. Events 3 and 4 must be completed before Event 5, but they consume less time than Event 2 on the critical path; Path 1-3-4-5 has one week of slack time. If the engineers reevaluated Process 4-5, for example, and decided it would take three rather than one week (and nothing else changed), then Path 1-3-4-5 would become the new critical path. Often, activity network diagrams are drawn to a time scale to provide a visual sense of the duration of each task and the impact of delays along the critical path day-by-day.

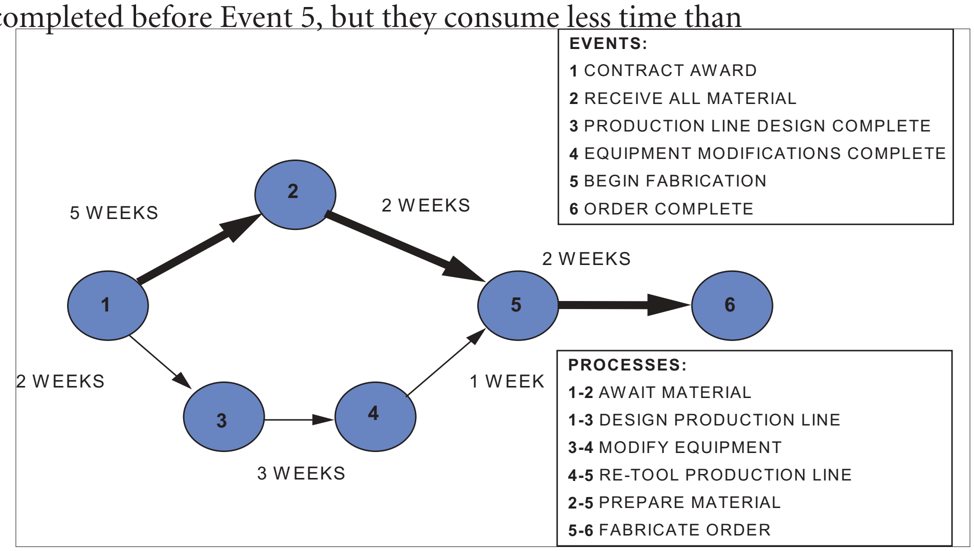


Figure 12-6. Product Activity Network Diagram.

Atlantic Undersea Test and Evaluation Center

PLAN of ACTION and MILESTONES

GOAL: Upgrade operational procedures, communications practices, range safety procedures, and training. Provide higher quality services to various range users while minimizing the risk of range safety violations.

OVERALL RESPONSIBILITY: LT D. P. Montague **APPROVED:** CDR M. WATERS

REVIEW DATE: ________________ **REVIEW DATE:** __________________

REVIEW DATE: ________________ **REVIEW DATE:** __________________

COMPLETION DATE:_______________ Page 1 of 4

ACTION STEPS/TASKS	START DATE	COMP DATE	PERSON(S) RESPONSIBLE	REMARKS
1. Establish written policy on roles, responsibilities, and interaction of RSO/CDO/RSWO				
2. Establish written policy on RSWO relief reports.				
3. Establish written policy on CDO and RSWO log keeping.				
4. Establish written policy on communication casualty procedures.				
5. Establish written policy on time allowed to process/transmit various types of messages.				

Figure 12-7. Service Improvement Plan of Action and Milestones.

Plans of Action and Milestones

A Plan of Action and Milestones lists all the steps that must be accomplished to implement a program or policy. It specifies by name or department who is responsible for each task and its due date, as shown in figure 12-7. The managers review and update the Plan of Action and Milestones periodically and maintain their portions continuously. The Plan of Action and Milestones may be based on a Gantt chart.

Verification

The purpose of verification is to ensure our implementation of the force planning alternative conforms to our expectations, guidance, and to regulations. Our expectations about the project started to form in the Definition Phase when we posed such questions as: What is the expected or required outcome from this decision? What different perspectives or opinions exist with respect to the problem? What are the key facts and assumptions?

During the Analysis Phase we identified criteria to choose between alternatives. These criteria are often reflected in formal documents from the Defense Acquisition System as Key Performance Parameters and by targets or policy goals. By the time the reconciliation process was complete, we codified additional expectations into the Acquisition Program Baseline and procurement specifications and validated or modified our policy goals.

Verification is the process of measuring the product and process and comparing the results to the expectations. This feedback may be used in a variety of ways, from making minor adjustments to system performance to termination of the entire project. As the Execution Phase progresses, verification ensures we meet our cost, schedule, and performance goals. Verification continues during the system's or policy's operational phase to monitor performance and identify improvements or changes to the system.

VERIFICATION MEASURES

The key to verification is deciding what to measure. For the hardware product or new policy, all attributes of the system are candidates for verification. The criteria we used in the analysis are excellent starting points for deciding what to measure. We use the same logic and tests for validity, reliability, and practicality while selecting what to measure. Verification measures with high validity compare the projected costs and schedules to what is actually happening. More frequent measurement intervals contribute to reliability and permit less dramatic adjustments and corrections. How often it is practical to measure depends upon the policy, type of system, the data measured, the urgency of conducting the measurements, and the penalties for making a mistake in implementation. It also depends on the accuracy and feedback requirements of the overseeing organization.

Once we decide what to measure, we must choose techniques with high levels of reliability and practicality to gather and track data. Some of the more common verification measurements are test programs, audits, sampling, exercises, and simulations:

- Formal test programs are an integral part of the acquisition of any major weapon system. The contractor will normally provide performance data as a contract deliverable. The potential for biases in contractor-provided data might necessitate the use of in-dependent testing agencies. The scope of such tests must be balanced against their cost and the perceived reliability of the contractor's data.
- Audits are systematic examinations of program plans and data to determine the efficiency and effectiveness of the implementation activities. They check for compliance with organizational procedures as well as accuracy and completeness of administrative records and reports. Finally, they ensure public funds and resources are properly protected and effectively used in achieving the system objectives.
- Statistical sampling is a practical way to gather data without actually observing the entire population. Sampling saves time and money. When done properly it is also very accurate and reliable. It is commonly used to verify a defense system's cost, schedule, and performance characteristics during development, test, and operational deployment. The data may be used to predict the system's ability to meet its objectives.
- Exercises test the system under operational combat-like conditions. Conducting regularly scheduled exercises provides the decision maker with important information on a system's ability to achieve the objectives over its life cycle.
- Simulation is used to test and predict system performance when real operational tests are not practical or when verification depends on some future uncontrollable event. War gaming is a type of simulation widely used in DoD.
- Surveys are the usual tool for verifying policy effectiveness just as they were used to choose a policy alternative in the first place; see Chapter 9, "Policy Analysis."

MEASUREMENT OBJECTIVITY

Once we know what and when to measure and how the data will be collected, we need to decide who will measure. To ensure objectivity, the organization that manages implementation should not be asked to perform verification. Each service has an Inspector General and there are several government audit agencies like the Congressional Budget Office and the General Accounting Office that have the professional and potentially unbiased ability to independently evaluate force planning implementation. These audit teams verify system or policy effectiveness using a variety of techniques such as inspections, testing, correlation analysis, simulated activities, and surveys. The services each have independent testing and evaluation commands that subject new weapons and support systems to exhaustive series of physical and simulator tests before the service resource managers, in a separate chain of command, can approve them for production.

Who receives these audit reports depends upon the program or policy and the service charged with execution. Clearly program managers and contractors have a strong and vested interest in an independent comparison of their hardware's actual performance to its actual performance. Many other stakeholders in the decision have an interest in test data. The other services are interested in the joint aspects of the project. Congress and comptrollers are interested in cost performance. A host of others, including other nation's governments, academics, businessmen, environmentalists, etc. all seek data that serves their interests and objectives.

Summary

Successful execution of a force planning alternative depends entirely on the earlier phases of the decision making process. The roots of implementation and verification begin with the problem definition and we embellish them in each subsequent phase, thus the direction implementation and verification will take is well set before we enter the Execution Phase.

Implementation is the process of shifting from choosing and reconciling a program or policy alternative to fielding an operational system or implementing a policy change. Implementation includes three principal activities: developing a plan to achieve the objective, organizing the resources to do the job, and managing the process until successful completion. There are a variety of powerful management tools available to achieve these ends.

Verification begins once implementation is underway. It ensures that system cost, schedule, and performance objectives are met. Verification uses technical measurements and professional judgment to compare the system's actual performance with targets and goals - our expectations. Organizations free from conflicts of interest, without bias or advocacy, provide oversight of implementation. The process of verification, if done properly in peace-time, will provide reasonable assurance of that our force planning objectives in peace will lead to success in war.

CHAPTER 13

AFTERWORD

… Reason and calm judgment, the qualities of a leader.
-Tacitus, 55-117 A.D. History

Leadership: The Power of Rationality

THE EXECUTIVE DECISION MAKING course taught by the U.S. Naval War College has three objectives. The first is to familiarize you with the context of many force planning and subsequent programming decisions: the formal U.S. defense resource allocation process. The second is to introduce you to the skills and tools you need to solve complex force planning and programming problems. The last is to acquaint you with ways of using those same skills and tools to reconcile differing views so that you can build consensus and execute a course of action favorable to your organization.

Your professional contribution to your service or agency and to the nation's defense will be increasingly measured by your ability to solve these problems and make decisions rationally; that is, to choose the course of action most likely to secure the nation's objectives within the available resources. Given the uncertainties of the future, the pace of change, and limited resources, this is no easy task. But there is a deeper, less tangible, and more fundamental reason to master the skills of rational problem solving and decision making: good leadership.

Leadership is the ability of an individual to move an organization or group toward an objective. As you know from your own experience and what you have learned studying at the U.S. Naval War College, leadership consists of many skills. The situation determines which of those skills is most important. In the context of a battlefield, the ability to inspire others is an especially important leadership skill. In the context of large, complex organizations, leaders are often characterized by how effectively they interact with bureaucracies and other organizations. In political contexts, the ability to persuade and advocate is an important part of leadership.

In our view, one characteristic underpins effective leadership in all situations: the ability to make rational decisions. Rational decisions offer the best chance of success for you and your organization. Sometimes making a rational decision means extended work with a computer. Sometimes it means reliance on experience and intuition to make an instantaneous decision. Sometimes it means minimizing risk. Sometimes it means waiting for uncertainty to diminish. Sometimes it means taking a deep breath and stepping into the dark. In each of these circumstances, however, effective leaders choose the course of action whose costs and benefits offer the best chance of success.

We think analysis is central to that effort. The most inspirational leaders will soon lose their effectiveness if they consistently lead their people and organization to failure. The most skillful bureaucrat is useless if his or her organizational victories do not enable the service or agency, and its people, to reach the proper objective. Beneath the bureaucratic politics and the inspiration of charismatic leadership there must be a course of action that the leader chose. In the end, the correctness of that choice determines the effectiveness of the leader. The skills and tools we have discussed in this course are central to making those choices. This is the power of rationality.

We are not suggesting that in the hurly-burly of organizational life, rationality is the only thing that matters or that the most rational decision always wins. But it is naive and incorrect to believe, as some do, that "it's all politics" or "it's all the budget." Arguments and debates that accompany all the important decisions about force planning are matters of substance. DoD may not always be able to choose the best course of action; politics, the budget, personalities, history, and chance all influence decision making. Clearly, irrational courses of action are seldom, if ever, among the alternatives senior leaders seriously consider, much less select. There is no doubt that budget constraints are a critical factor in defense decision making in the current era. In DoD, our annual Fiscal Guidance tells us only *how much* we can spend. It never tells us on *what* to spend it. Tight as it may be, we may spend the defense budget, approximately $345 billion per annum, in a myriad of different ways, each with different consequences. To say "it's all budget" is to say that the size of the budget alone completely determines what is in the budget. That is plainly untrue.

It can be difficult to see the importance of rational decision making while in the midst of the Defense Department's formal resource allocation process. Indeed, as we have discussed, there are numerous ways rational decision making may be derailed by malfunctions in that process. For example, if the Defense Planning Guidance is too late or too general, the service force programmers use proxies for guidance, such as the inertia and projections of last year's Future Years Defense Program. This is a form of decision making by procedure, and it is one explanation for the slowness of change in the Pentagon. Similarly, if the Future Years Defense Program arrives at the DoD Comptroller exceeding the President's projected budget, then time constraints may force the Comptroller, rather than senior defense planners and strategists, to make crucial force planning decisions by cutting programs. Such events subordinate rational decision making by senior DoD leaders.

Keep in mind that the formal defense resource allocation process was designed to institutionalize rational decision making. The extent to which we achieve this objective in DoD varies year by year, depending upon the circumstances and the personalities involved. Most participants, researchers, and observers grade the process as adequately rational in a bad year and better in a good one. This can be hard to see when you are in the thick of it, in the same way that a well-organized battle seems chaotic at the foxhole level. By "adequately rational" we mean that the process usually does an acceptable job of developing strategic objectives and selecting courses of action that are likely to achieve those objectives without excessive risk. There is no reason we should be satisfied with this level of performance, and many people are not. We can expect to see continuing changes in the formal process as, every few years, high-level panels are appointed to revise one or another feature of the Joint Strategic Planning System, the Planning, Programming, and Budgeting System, or the Defense Acquisition System.

The process does succeed at producing a rough correspondence between ends and means; in other words, adequately rational resource allocation. This does not happen by accident. Nor could it happen at all if only politics or the budget explained defense resource allocation decisions. This result is obtained year after year because there is enough rational decision making embedded in the formal process to produce it, even though, day-by-day, it may not be easy to see.

Finally, wherever your future assignments take you, we believe that the most important contribution this course makes is a personal one: the intellectual habit of defining a problem carefully, developing realistic alternatives, thinking clearly and objectively about their comparative strengths and weaknesses, and reconciling your choice with that of other parties involved with the same problem. That way of solving problems, large and small, professional and personal, we call the discipline of critical thinking. In your professional and personal lives, proponents of ideas and products who seek to persuade you to take the course of action they recommend surround you. These people are usually good at their jobs. They can muster powerful facts and arguments for alternatives that, in the end, may not be productive. They may be completely convinced of their own correctness. Yet every day, you have to decide who is making sense and who is not, who is telling you the truth and who is not. The discipline of critical thinking taught in this course is the most effective way we know to weigh their arguments and rationally choose among them. In our view this skill, the ability to make sense of confusion and to "see" clearly, is a crucial part of your ability to lead in the future.

CASE STUDY: AFTERWORD
USMC MEDIUM-LIFT REQUIREMENTS: THE V-22 OSPREY AND HELICOPTERS

Throughout 1991 to 2002, funding for the V-22 Osprey Program has been stable. The scheduled procurement of 360 V-22s for the Marine Corps, the same number recommended by the 1997 Quadrennial Defense Review (and remarkably close to the Institute for Defense Analyses's recommendation), 50 Special Operations V-22s for the Air Force, and 48 Combat Search and Rescue V-22s for the Navy is still on-track. Bell-Boeing-Textron, Inc., built four flying V-22 prototypes and ten Low-Rate Initial Production V-22s. However, two of the prototypes and two of the production aircraft have crashed, altogether killing 30 Marines. The most recent incident, as of this writing, occurred December 5, 2000, immediately before the Defense Acquisition Board was scheduled to consider authorizing full production. Then-Secretary of Defense Cohen postponed that decision until several investigations were completed. These accidents re-opened the debate in Congress, the Pentagon, and in the defense community concerning whether the Osprey Program should continue.

In the wake of the December 2000 crash, Secretary of Defense Cohen appointed a four-person panel to make recommendations to the new administration on the V-22 Program. The members examined the V-22 Program's training; aircraft engineering and design; production and quality control; operational suitability; and flight safety and performance. The panel reported their findings to Secretary of Defense Rumsfeld in April 2001.[1] Additionally, Senator John Warner of the Senate Armed Services Committee held hearings on the V-22.

The contentious issues, as in 1990, still include cost and effectiveness but now, with many hours of flying the aircraft available for evaluation and analysis, and the findings of the panel noted above, there is much more discussion of risk. The V-22 has largely met its cost goals and projections; at $40B for the overall program, production model Ospreys will cost $83M each, including research and development costs. But, as we discussed in Chapter 4, the important cost for decision makers is the relevant or production model cost of each V-22 from here on: approximately $44M per aircraft. The sunk costs of the V-22 Program, by reducing its relevant costs remaining, are making a replacement helicopter program increasingly unattractive, unless that helicopter comes off the shelf, like the $8M UH-60 Black-hawk.

The V-22 has another important edge over helicopter program alternatives that grows larger with the passage of time: the Marines need V-22s to replace CH-46s as soon as possible and the Air Force needs to phase out the MH-53 Pave Low in 2007. The CH-46 already operates under severe flight restrictions and MH-53 missions require extensive risk assessment. Both aircraft require intensive maintenance; their cost per flying hour is steeply increasing. Because both services need production aircraft in the near term, there is precious little time to begin a new medium-lift program from scratch.

The effectiveness issues remain similar to those in the Institute for Defense Analyses study we examined earlier: Is the increased speed and endurance worth the additional cost? The background, approaches, models, and results are likely to be similar, too. The V-22, if it works as designed, performs the missions better than helicopters. Is its superior performance worth the cost? More than likely, the answer will again be yes, as in the previous seven studies.

What is different now is the opportunity and requirement for more strenuous risk assessment, in several dimensions. How willing are Congress and DoD to accept risk in the form of reduced readiness and peacetime casualties to achieve superior operational performance? First, there is the question of the vulnerability of tilt-rotor technology itself, especially under combat conditions, which could change our perceptions of its effectiveness. Second, the models must be able to show the benefits of new employment opportunities for the Osprey compared to historical helicopter operations.

The V-22, as a hybrid aircraft, neither glides well with its small wing area, nor do its tilt-rotors generate enough lift to auto-rotate downward as a helicopter can if it loses power. Because it cannot fly with a single tilt-rotor, powered or otherwise, it is susceptible to catastrophic failure if either rotor is damaged. (A fire in an engine nacelle caused one of the prototypes to crash in June 1991.) Drive trains run through the length of both wings so that one engine can power both tilt-rotors, however the wings themselves are necessarily unarmored and vulnerable. The new analyses should assess the Osprey's vulnerability in combat scenarios, e.g., assaults and extractions under fire, in more detail than IDA was able to complete. The analysis should include the probability of being hit and the effects of battle damage if hit, vis-à-vis helicopters.

According to the accident review board, the April 2000 crash was due to a combination of human-controlled factors, principally low forward air speed and a high rate of descent, in excess of 2,000 feet per minute. At high rates of descent, the rotors of a helicopter or an Osprey lose their ability to create enough lift in the turbulent air, a condition called vortex ring state. To re-

1. John R. Guardino, "MV-22 Osprey Reeling From Latest Disclosures, Media Attacks," *Helicopter News*, January 21, 2001: 1.

cover flight control, the pilot increases forward speed to punch through the turbulent air into clear air; this requires clear air ahead of the aircraft and sufficient altitude while the descent continues. The Osprey was, at the time of the crash, at low altitude making a landing approach behind another V-22 (creating turbulent air ahead of it), so there was little or no opportunity for the pilot to recover control. The V-22 was limited to 900 feet-per-minute descents at the time of the accident; ironically, after additional flight testing, the Naval Air Systems Command expanded the V-22 flight envelope to allow descents up to 1400 feet per minute.[2] Another concern, unique to tilt-rotors, surfaced during testing: the possibility that each tilt-rotor may experience different degrees of vortex ring state and develop a lateral torque that would tip the aircraft over sideways.

The V-22 hydraulic system design incorporates technological risk and has been another area of many skeptics' concern. The Osprey hydraulic system operates under 5,000 psi, significantly higher than 2000-3000 psi found in most aircraft and helicopters. Designers selected the higher levels to reduce the size and weight of its components, but that has made them more prone to leaks and they leak more fluid faster. To compensate for the increased vulnerability, the Osprey's critical flight control systems are triple-redundant, i.e., there are three paths to get hydraulic fluid to essential flight controls. If the flight control computers sense a failure in the primary system, they open and close remote valves to change the path to the control surface actuators within 0.3 seconds. While rotating one of the tilt-rotors, the primary hydraulic system failed—a hose rubbed through, according to Lt. Gen. Fred McCorkle, USMC, then Deputy Commandant for Marine Aviation—and the flight control software failed, too, causing the December 2000 crash.[3] Unlike other aircraft that have non-hydraulic back-up systems for critical flight controls, e.g., high-pressure air flasks or electrical servo-motors, the V-22 must rely on hydraulics for tilt-rotor nacelle rotation because there is no alternative motive force strong enough to rotate them in flight.

The V-22 hydraulic system is maintenance-intensive, more so than the manufacturers indicated when they provided reliability estimates to IDA. The flight availability of the Osprey has been low, to the point where the training squadron (VMMT-204) commanding officer decided to misreport aircraft availability, and was relieved for doing so. Furthermore, the media has accused senior Marine aviators of publicizing rosy availability numbers to advocate a Defense Acquisition Board decision for full production. (None of the alleged false reporting is related to any of the accidents, i.e., no one recorded maintenance actions that were not actually completed.)

Viewed dispassionately, much of the confusion about availability and readiness reporting arose because there are three separate maintenance reporting methodologies within the squadron, each with different reliability problems. The manual system is based primarily on the mechanics' subjective evaluation of whether an aircraft is down (unavailable), mission capable (can fly at least one of its missions) or fully mission capable (can fly any of its missions). The second and third reporting methodologies are an older and an upgraded version of the Department of the Navy's automated systems that report readiness based on maintenance actions. According to Marine Corps spokesman Lt. David Nevers, the upgraded automated system is the most stringent of the three and, while it is least subject to manipulation, it also produces distortions like reporting an aircraft down that is undergoing a visual inspection because an access door is open. As a result, for the month of November 2000, the three systems reported mission capable rates 73 percent, 57 percent, and 27 percent, respectively.[4]

2. John R. Guardino, "Catch-22 For the V-22," *Rotor & Wing*, February 2001.
3. Robert Wall, "V-22 Support Fades Amid Accidents, Accusations, Probes," *Aviation Week & Space Technology*, January 29, 2001: 28.
4. Lisa Troshinsky, "Corps Says V-22 Readiness Confusion Caused by Using One System For OPEVAL, Another For Press," *Navy News and Undersea Technology*, Feb.5, 2001:1.

The readiness issue plays into another area of risk that Congress and the Pentagon need to explore: the risk to the production schedule. If the V-22 falls significantly behind schedule, Marines and Special Operations Forces incur greater personal and mission risk by pushing their aged helicopters past their already extended service lives. But the DoD Inspector General, the General Accounting Office, and the DoD Director of Operational Test and Evaluation have all criticized the V-22 program managers for curtailing tests to keep the V-22 on schedule, despite significant technical deficiencies. All three recommended slowing the program down to resolve technical problems.

The DoD Inspector General's August 2,000 report identified 22 major, documented deficiencies that the Department of the Navy was going to waive to get the V-22 into production: "Program officials accepted a higher level of risk to get the program into production, despite uncertainties that the system would work as intended, rather than delaying the program and risk losing the funding." As a result of the report, the Pentagon postponed the production decision to December 2000.[5]

After an eight month evaluation, in November 2,000, Mr. Philip Coyle, then DoD Director of Operational Test and Evaluation, reported to the Secretary of Defense that his organization identified 177 failures in flight-critical systems among 723 other malfunctions including safety-related failures that could trap air crews inside the aircraft or cause in-flight fires. Nineteen test criteria were waived from evaluation, including: shipboard operations; rescue hoist and fast-rope operations; flight performance operations under icing conditions; and aerial combat maneuvering.[6] (Since November, some these tests have been completed.) Mr. Coyle declared the Osprey was not operationally suitable because of reliability and maintenance concerns and therefore not ready for production.

Also in November 2000, the General Accounting Office, relying at least in part upon Mr. Coyle's evaluations and the investigation of the Osprey crash in April 2000, described a decision to approve full-scale production as fraught with "significant risk" because the "baseline development flight test program [was] restructured numerous times to meet program cost and schedule pressures…. Knowledge of V-22 design and performance parameters falls short of what should have been known before beginning production…. Developmental flight testing was deleted, deferred, or simulated. Operational test waivers and limitations reduced testing for operational realism."[7]

But the Marines' need to get replacement helicopters to the operating forces as quickly as possible is compelling; some critics felt that this was compromising Marine leaders' rational decision making. Based upon the publicly reported transcript of his comments, the VMMT-204 Commanding Officer wanted his Marines to shade the aircraft availability reports specifically to get the V-22 past the Defense Acquisition Board's milestone decision to approve full production, despite the warnings in the Pentagon and GAO reports. To remove concerns that the Marine Corps has been seduced by the V-22 program, the Commandant shifted the investigation about the misleading readiness figures, and the possibility of improper command influence on the VMMT-204 Commanding Officer, to the Department of Defense Inspector General.

In a historical context, the V-22's developmental track record of casualties and mishaps in its first five years is not very different than other rotary-wing and some fixed-wing aircraft that intro-

5. Dan Hardy and Ralph Vigoda, "V-22 Osprey Has Strong Allies, Doubters," *Philadelphia Inquirer*, December 14, 2000.
6. Elaine M. Grossman, "Pentagon Test Director Found 177 Osprey Failures Endangered Safety," InsideDefense.com, February 8, 2001.
7. Christian Lowe, "Navy Cut Osprey Tests That Could've Shown Fatal Flaw," *Defense Week,* January 29, 2001: 1.

duced new technology. For example, according to Naval Safety Center data for Class A mishaps (loss of life or greater than $1M of damage), the CH-53D heavy-lift helicopter had nine Class A mishaps in its first five years, the H-3 Sea King helicopter had 28, the UH-1 Huey helicopter had 43, and the F-14 variable geometry wing interceptor had 27.[8] Because naval aviation safety has improved dramatically in recent decades and our tolerance of casualties has diminished, we may be holding the Osprey to an unrealistically high standard.

In addition to the lack of an obvious helicopter alternative that meets the Marines' requirements, there is another down side to canceling the V-22. President Bush campaigned on a promise to strengthen the U.S. military, in part by skipping a generation of technology. Presumably, the V-22 is exactly the kind of next-generational technology he believes is important for the new security environment. The future of the civilian application of tilt-rotor aviation, the BA609, is tied closely to the fate of the Osprey. By March 1999, Textron had 37 customer commitments for its civilian model—which has a 3000-psi hydraulic system.

The Marines are standing by the Osprey because they believe tilt-rotor science is a sound application of achievable technology and the V-22's performance is essential to executing their Operational Maneuver From The Sea concept. On The NewsHour with Jim Lehrer on January 22, 2001, Marine Commandant Gen. James Jones said, "The technology is not the issue, as far as the accidents… I'm confident in the technology. I'm confident in the research that's gone into it. I'm confident in the people who advise me with regard to the potential of this airplane, but we are not going to do anything reckless."

As of this writing the V-22 is undergoing a complete program restructure and a major engineering redesign and modification. To keep the program alive, 11 aircraft a year are being produced. Additionally, the Marine Corps is now committed to an event, vice schedule, driven schedule. Accepting a two-year program pause, the V-22 is now anticipated to be ready for a production decision in August 2003.

✶ ✶ ✶ ✶ ✶ ✶ ✶

As this drama further unfolds, we invite you to use the Executive Decision-Making Frame-work to evaluate the problem definitions, studies, decisions, and reconciliation about the V-22 Program that ensue. Observe how various stakeholders define the problem, especially when they define it with different emphases on cost, effectiveness, or risk. Notice, too, the role analysis plays in shaping upcoming decisions—what criteria would you use to evaluate the risks associated with tilt-rotor technology or the Osprey hydraulic system? Would you be willing to preserve effectiveness and reduce risk, despite the increased cost and the effects of delays in production, to slow the program down, or can these problems be fixed in stride? How much subjectivity should the Naval Air Systems Command allow operators to evaluate aircraft readiness and availability?

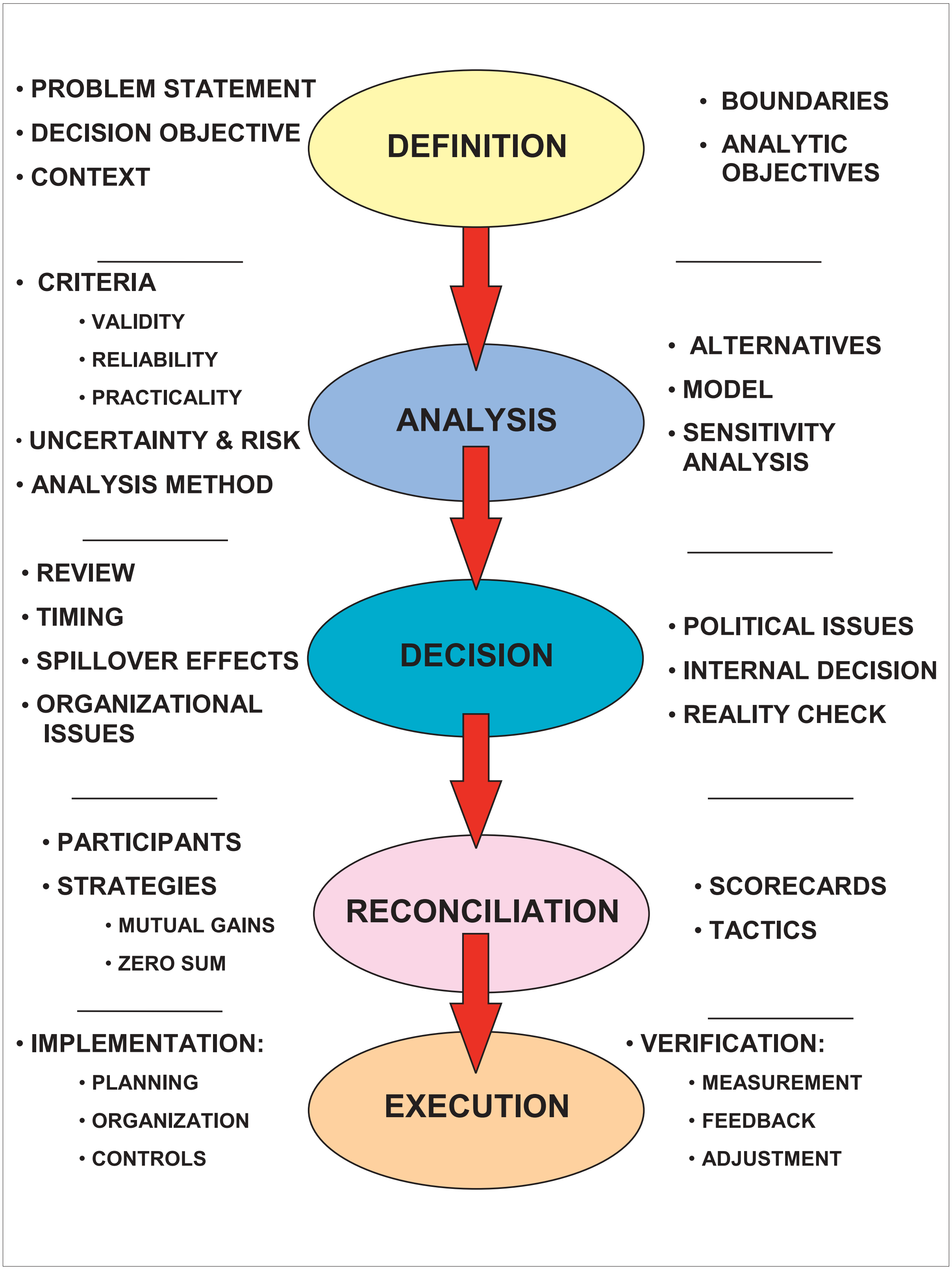
Examine, too, how executive decision makers combine experience and analysis to reach rational decisions—rational at least from their perspective. After leaders make their decisions, identify their approaches to reconciling their decision with other stakeholders, how they use analysis to bolster their arguments, and whether they use a mutual gains approach or more traditional zero-sum strategies.

With apologies to Shakespeare, all the Pentagon is about resource allocation, and all the men and women merely players; we have our exits and entrances; and one man in his time plays many parts as advocate and adjudicator in the course of many executive decisions.

8. Stratfor.com (February 6, 2001), http://ebird.dtic.mil/Feb2001/s20010208fate.com

APPENDIX 1

AN EXECUTIVE DECISION-MAKING FRAMEWORK



APPENDIX 2

AN EXPANDED EXECUTIVE DECISION-MAKING FRAMEWORK

DEFINITION

IDENTIFY THE PROBLEM. Is there a problem and do we need to solve it now? Have the decision maker approve the major elements of our problem definition.

- Some problems do not need to be solved, and some problems cannot be solved. Some problems can be tolerated indefinitely and others will disappear by themselves.
- Identify the importance, urgency, and magnitude of the problem to determine the resources our organization will devote to solving this problem.
- Is this problem part of a larger problem? Study the backdrop of the problem and how it is related to other problems.
- Can this complex problem be deconstructed to make it easier to work on? Organize the sub-problems hierarchically, show their links to one another, or their sequence in relation to each other.

CRAFT A PROBLEM STATEMENT. Describe the requirement we have that is not being met.

- What is our desired end state?
- What is our present observed condition? Is our equipment, organization, doctrine, or policy inadequate or nonexistent?

SPECIFY THE DECISION OBJECTIVE. What is the desired outcome of our organization's decision making?

CONSIDER THE PROBLEM CONTEXT. Brainstorm an exhaustive list of the factors bearing on the problem.

- Who are the stakeholders? What are their concerns? Will we consider their positions before or after we choose our organization's preferred alternative?
- What triggered our present decision making? Why was this problem brought before us now? How urgent is a solution?
- List the information we already have as influences. Decide whether the items we have brainstormed are internal, to be considered now, while choosing an alternative, or whether they are external to our decision making, meaning we deal with them later during reconciliation.

SET THE PROBLEM BOUNDARIES. Limit the scope of the problem-solving effort to conform to the resources our organization is willing to consume, and identify factors that are immutable.

- Establish the timeframe we have available for studying solutions and the planning horizon we desire our solution to span: a quick fix or a long-term solution?

- Identify our organization's rule sets that bear on this problem. Are some alternatives off limits?
- List facts that are already known, and the assumptions defined or given by the decision maker that are non-negotiable.
- Make assumptions, the statements we take to be true without proof, in order to cope with uncertainty. They must be necessary to proceed and they may cause controversy later; therefore, we should make as few assumptions as possible.

SPECIFY THE ANALYTIC OBJECTIVE(S). Identify the objective of each analysis that will support the decision.

- Analytic objectives are subordinate to the decision objective.
- Each analytic objective must be capable of independent analysis.
- Complex decision will require several supporting analyses, each with a separate analytic objective.

ANALYSIS

ESTABLISH CRITERIA. All programs and policy alternatives can be described in terms of attributes; we call those attributes that will help us opt for an alternative Criteria. We choose criteria that highlight differences between alternatives in effectiveness, cost, schedule, risk and uncertainty where those differences exist. Good criteria have the following characteristics:

- Direct connection to the analytic objective
- Inclusiveness
- Precise definition
- Measurability
- Uniqueness

EVALUATE THE CRITERIA. Examine the criteria, individually and as a set, for:

- Validity: Are we measuring the right things?
- Reliability: Are we measuring well, with the right level of fidelity? How large are our measurement errors, and are our results consistent?
- Practicality: Are the resources we consume worth the knowledge we gain?

DETERMINE THE ANALYTIC METHOD. There are three basic methodologies:

- Exploratory: Wide-ranging, few restrictions, softer data, used for most mission needs identification and concept development.
- Cost-Risk-Effectiveness: Highly structured, data intensive, often using a systems approach, used to compare alternatives.
- Causal: Used in policy analysis to identify cause and effect relationships.

EVALUATE RISK AND UNCERTAINTY

- Identify objective and subjective probabilities and their importance to the decision.
- Decide where to reduce risk or uncertainty by doing additional research, buying it out, or generating expected values.
- Improve subjective probabilities by gathering more expert information.

DEVELOP AN ANALYSIS PLAN. The defense executives and analysts review the analysts' proposals to achieve the analytic objective(s).

CONSTRUCT OR IDENTIFY THE ALTERNATIVES. Sometimes we know the alternatives before we set out to solve the problem; sometimes we identify them as part of the process. Where this step occurs depends on the decision objective and the analytic method. We desire a set of alternatives with:

- Breadth: The set should span the range of solutions.
- Viability: Each option must meet our minimum requirements (or be modifiable to meet them).
- Neutrality: Each option should be presented without bias.

ORGANIZE THE CRITERIA: MODELING. Models are simplifications of reality that enable the decision maker to better understand the differences among alternatives. We select between analytic, force-on-force, and policy models depending upon the decision. When we choose a model, we consider its level of:

- Abstraction: How closely does it mimic the real world?
- Predictability: How well does it forecast the alternatives' behavior?
- Complexity: How difficult is it to build, understand, use, and explain? What does it leave out?

EVALUATE THE MODEL. Evaluate the model in terms of:

- Validity: Does it capture the most important aspect of the alternatives' behavior? What underlying assumptions or critical processes are missing from the model?
- Reliability: Is the model internally consistent and can we rely on its predictions?
- Practicality: Will additional resources improve the usefulness of the model?

EVALUATE THE ALTERNATIVES USING THE MODEL. The outcome of the model should be more than a rank ordering of the alternatives; the results should demonstrate how the alternatives differ. Can lower scoring alternatives be modified to improve their preference with respect to the decision maker?

CONDUCT SENSITIVITY ANALYSIS. Examine the degree to which important assumptions, weighting factors, or variables can be adjusted to alter the outcome.

DECISION

PREPARE TO DECIDE. Adapt the analytical results by assessing:

- Changes in problem definition
- Organizational issues
- Political issues
- Spillover effects: What side effects are likely to be generated by each alternative?
- Timing: Do we need to decide now and what are the penalties for postponing the decision?
- Display the results of the analysis for the Decision Maker.

CHOOSE THE BEST ALTERNATIVE FOR OUR ORGANIZATION. Starting with the preference indicated by the analysis, combined with military judgment and organizational considerations, identify the best option for our organization.

DO A REALITY CHECK. Is this option feasible, i.e., legally, ethically, organizationally, politically, and economically acceptable?

PRESENT THE RESULTS. Prepare the briefing or report that documents the decision.

RECONCILIATION

IDENTIFY PARTICIPANTS. Identify the other principal participants in the decision process and their positions and interests; also identify others who are affected by the decision but who may not participate in making it.

DETERMINE A RECONCILIATION STRATEGY. Decide on the overall approach that our organization will take to protect and advance our interests.

- Identify opportunities for mutual gains and value adding.
- Identify situations where zero sum approaches may be necessary.

MAKE SCORECARDS. Include the minimum acceptable outcome, best alternative to a negotiated agreement, most likely outcome, and best outcome for our organization and each other participant.

NEGOTIATE. Use the following tactics:

- Separate the people from the problem.
- Focus on interests not positions.
- Promote confidence-building.
- Invent options for mutual gain and add value, e.g., consider linking this problem to another issue.
- Insist on objective criteria.
- Husband concessions.
- Break deadlocks.

EXECUTION

IMPLEMENTATION. Implementation planning should take advantage of our earlier work; e.g., our boundaries may provide guidelines for implementation processes. Select management aids that best fit this project.

- Plan: Set a schedule and decide upon a sequence of events that will implement the alternative.
- Organize: Decide who will be responsible for the different aspects of execution and delegate authority to them.
- Control: Identify measurement methods, timing, and standards.

VERIFICATION. The verification process must be in place when implementation begins. It is the process of measuring the product and implementation process and comparing the results to expectations.

- Measures: Decide what to measure; our criteria may be helpful.
- Techniques: Establish how and when data will be collected.
- Objectivity: Decide who will measure (to ensure reports are unbiased).

APPENDIX 3

TESTIMONY FROM THE SENATE HEARING REGARDING THE V-22 OSPREY

THIS IS THE RECORD OF DR. L. DEAN SIMMONS'S and Dr. David Chu's testimony about the strengths and weaknesses of IDA's study and several prominent Senators' reactions to DoD's insistence on canceling the V-22 program. Dr. Simmons led the Institute for Defense Analyses study that we used at the end of many chapters in the text to illustrate the Executive Decision-Making Framework. Dr. Chu, the Director of Program Analysis and Evaluation of the Office of the Secretary of Defense testified after Dr. Simmons. Dr. Chu presented Secretary of Defense Cheney's position that the V-22 was not affordable. He received a heated reaction from several Senators. This is a rare example of two senior analysts placing opposing views on the record. This hearing also underscores the importance of supporting DoD leaders with high-quality rational analysis, captures the essence of the often fiery V-22 debate, and it shows the spirited nature Congressional inquiry can take when major defense programs affect jobs in members' districts.

INSTITUTE FOR DEFENSE ANALYSES STUDY OF
THE V-22 OSPREY
THURSDAY, JULY 19,1990

U.S. SENATE,
SUBCOMMITTEE ON DEFENSE.
COMMITTEE ON APPROPRIATIONS,
Washington, DC.

The subcommittee met at 9 a.m., in room SD-192, Dirksen Senate Office Building, Hon. Daniel K. Inouye (chairman) presiding.

Present: Senators Inouye, Bumpers, Lautenberg, Stevens, Garn, Kasten, D'Amato, Specter, and Gramm.

INSTITUTE FOR DEFENSE ANALYSES

STATEMENT OF DR. L. DEAN SIMMONS, RESEARCH STAFF, SYSTEM EVALUATION DIVISION, INSTITUTE FOR DEFENSE ANALYSES

ACCOMPANIED BY DR. DAVID L. RANDALL, DIRECTOR, SYSTEM EVALUATION DIVISION, INSTITUTE FOR DEFENSE ANALYSES

OPENING REMARKS OF SENATOR INOUYE

Senator INOUYE. The subcommittee meets this morning to discuss the future of Marine Corps aviation during a time of increasing fiscal constraints on defense spending. Today's hearing focuses on the V-22 tilt rotor aircraft, a program which has been killed by the Secretary of Defense but resurrected by Congress.

It is an understatement to say that the Secretary's decision to terminate the V-22 has been controversial. The issues are complex and the decisions Congress will make on the in-vestment or savings amount to billions of dollars. The V-22 proponents say that the aircraft represents a revolutionary technology, a technology with military and civilian aviation bene-fits, which justify the program costs. Opponents suggest that these benefits are not commensurate with the amount of scarce resources the V-22 would divert from the other higher priority defense programs.

The Marine Corps clearly wants the V-22 despite its official position to support the President's budget. The Defense Secretary is just as adamant in his position to terminate the program. So we are holding this hearing this morning to shed some light on both sides of this debate. The witnesses today are the Honorable David Chu, Assistant Secretary of Defense for Program Analysis and Evaluation, and Dr. L. Dean Simmons of the Institute for Defense Analyses. Dr. Simmons was the project leader for an IDA special study entitled "Assessment of Alternatives for the V-22 Assault Aircraft Program." Gentlemen, we look forward to your testimony. After any opening remarks by members of this subcommittee we shall first hear from Dr. Simmons. Dr. Chu will testify after that. Senator Specter.

STATEMENT OF SENATOR SPECTER

Senator SPECTER. Thank you very much Mr. Chairman. To begin with, Mr. Chairman, I thank you for convening this special hearing. I believe it to be necessary as the testimony presented by the military has very forcefully supported the V-22 Osprey. General Gray, head of the Marine Corps, testified as to some of the objections and called them totally ridiculous.

An independent report by the Institute for Defense Analyses has categorically stated that the V-22 Osprey is vastly superior to any alternative, and it is well within the budget constraints. The Department of Defense early opposed the Osprey, and has candidly maintained this position despite an overwhelming case in its favor.

The central point of our inquiry here today is to focus on the facts. What can the Osprey do?

Mr. Chairman, I would suggest that there is a very important subsidiary question in terms of procedures and, really, the good faith of the Department of Defense in terms of opening this process to Congress.

Last year, both houses of Congress mandated an independent study: a study to deter-mine facts. As the process has gone forward, there have been reports of a preliminary conclusion not made available to Congress. There were meetings on April 16 with no congressional participation. There was strong insistence in the Congress on getting access to the re-port which was finally made available under strong pressure by this Senator, talking to Dr. Chu and finally to the Secretary of Defense, on Friday when we left for the Fourth of July recess, which has the unmistakable imprint of trying to delay it as long as possible until 535 Members of Congress have left town.

The central question here is the quality of the plane. I think that there is absolutely no basis for any contention that this is a parochial issue. Obviously, Senators from Pennsylvania and Texas are going to take a special look because our states are so heavily involved, but this is a matter of national defense. The very distinguished chairman of this subcommittee has been very forceful on this issue, and all it can do is fly to Hawaii.

The ranking member of this subcommittee said last year that he would not vote for an appropriations bill that did not contain the V-22 Osprey. Now what is a Pennsylvania Senator supposed to do? Absent himself from these proceedings? This is not a matter of Pennsylvania or Texas or parochialism. This is a matter of national defense. We know that we are now going into an era in which B-2s and long-range missiles are less important. We know we have problems in Panama. We know we have problems in the Persian Gulf. We know we need flexibility.

I reviewed the matter again yesterday afternoon with General Pittman, who originally opposed the V-22. Now he calls it absolutely indispensable. He says we can document that had we had the V-22 in Panama, we would have saved lives. Why save lives? Because the V-22 can land like a helicopter after flying long distances as a fixed-wing plane. You do not have to jump people on parachutes. There is nothing more important than saving lives when we ask our military personnel to take action in the national interest.

If I sound just a little bit perturbed, Mr. Chairman, it is because I am. I think it is really important to have Dr. Chu and Dr. Simmons head to head addressing the facts. All we want are the facts and let the chips fall where they may. I am satisfied after reviewing a 1,200-page report that the facts are overwhelmingly in support of the V-22, and that any fair-minded interpretation or conclusion will support the V-22.

Thank you, Mr. Chairman.

Senator INOUYE. Thank you very much. Senator Specter.

Now, Dr. Simmons.

SYNOPSIS OF STATEMENT

Dr. SIMMONS. Mr. Chairman, members of the committee, I am quite pleased to appear before you this morning in connection with IDA's assessment of alternatives for the V-22 assault aircraft program. The Institute is one of our federally funded research and development centers.

We carry out studies and analyses for the Office of the Secretary of Defense, the Joint Staff, the unified commands, and the Defense Agencies.

ASSESSMENT OF ALTERNATIVES

The V-22 assessment that I will be reporting on this morning was conducted at the request of the Office of the Secretary of Defense, which, as you know, had responsibility for the study that was directed by the Congress last year.

The detailed results of our study have been documented in a five-volume draft report. Copies of that report have been forwarded by the Secretary of Defense to the Senate Appropriations and Armed Services Committees and to the corresponding committees of the House of Representatives. The report is now being reviewed within the Department of Defense by the steering group established for our study, and after consideration of comments that they may provide, IDA will publish a final version of the study.

Later this morning I will be presenting a short briefing of our findings. In my opening statement I would like to discuss two of the key considerations that relate to the cost of the aircraft: the overall cost-effectiveness of the alternative fleets measured over a 20-year period that we used as the principal basis for our conclusions, and the difference in the near-term costs for the alternative fleets.

PROCESS LEADING TO STUDY RESULTS

Before doing that, though, I would like to take a few minutes to describe the process that we used at IDA to arrive at our study results. To conduct a study of this scope, we need access to a considerable amount of information, detailed data related to the performance and costs of the alternative aircraft, descriptions of how our various military forces would employ the aircraft in combat situations, and estimates of the performance of the weapons that might be used against the aircraft by potential wartime enemies.

To obtain this information we met frequently with the three military services that have proposed to use the V-22 Osprey and the assault helicopters: the Marine Corps, the Navy, and the Air Force. We also had frequent and detailed discussions with the contractors that would build the different aircraft Bell, Boeing, and Sikorsky in the United States and Aerospatiale, Augusta, and Westland in Europe. We had numerous discussions with representatives from our intelligence community to obtain information on potential threat systems.

As one might expect, the study has been subject to considerable oversight and review both by IDA and by the Department of Defense. As is the case with every major study conducted at IDA, a review committee was established by IDA to provide guidance and rigorous review of the technical and operational inputs and the analytical methods. The review committee was composed of knowledgeable individuals from outside IDA and appropriate members of IDA's management and research staffs. The committee was chaired by Dr. David Randall, Director of IDA's System Evaluation Division, which is the component of IDA where I work. Dr. Randall has accompanied me today and will be available to help answer any questions that you may have.

STEERING COMMITTEE

The Department of Defense provided guidance and direction to the study through a steering committee that included representatives from a number of OSD staff elements, in particular those responsible for program analysis and evaluation, for overseeing the development and acquisition of tactical combat systems, for tests and evaluation, and for estimating the costs of future military systems.

In addition, the steering committee included representatives from the Joint Staff and the Department of the Navy.

Between the time the study was initiated in September 1989 and April of this year, members of the IDA study team met with the steering committee five times to inform them of our plans and to update them on our progress. The committee, for its part, provided us with additional sources of information, occasionally suggested alternative approaches from the ones we had proposed and identified a wide range of sensitivities to be examined.

The steering committee also asked other Department of Defense agencies to review specific portions of the assessment. The amphibious assault scenarios and concepts of operation were reviewed by the Marine Corps War Fighting Center at Quantico and by the Joint Staff. The specific aircraft configurations used in the assessment and the flight performance projected for those aircraft were reviewed in detail by the appropriate staff elements at the Naval Air Systems Command. The development, procurement, and operating costs estimated for the aircraft were reviewed by other staff sections at the Naval Air Systems Command and by OSD's Cost Analysis Improvement Group.

Our last presentation to the steering committee took place on April 16, when we briefed them on the overall results emerging from our study. At that time, however, we still had some additional analyses to perform, some additional sensitivities suggested by the steering committee, and additional analyses that had been suggested by our own IDA reviewers, and we had not yet completed our written report. Those tasks were accomplished over the next two-months. Following IDA's detailed review, the draft study report was forwarded to OSD in late June and thereafter to the Congress.

COST ISSUES

At this time, I would like to address the two cost issues that I mentioned in the opening. In our study, we specifically structured the alternative assault aircraft fleets so that the cost to develop, purchase, and operate those fleets for 20 years would be the same in constant, fiscal year 1988 dollars. This so-called equal cost approach is frequently employed in assessments of this type because it allows easier comparison of the alternatives with respect to the missions to be performed. When all of the alternatives cost the same, the one that provides the most capability is the most cost-effective. This simplifies the identification of the preferred alternative over the 20-year timeframe examined.

As you are all aware, the V-22 Osprey fared quite well in this assessment. As you also know, there is another perspective regarding costs that the Department of Defense and the Congress must consider, particularly when resources are as tightly constrained as they are at the current time. The alternative aircraft programs, while equal in cost over 20 years, would require different levels of funding in the near term.

Because the V-22 would cost more per aircraft than any of the other alternatives, and because completing its development and setting up projection facilities would require ap-proximately $1 billion, the V-22 program would cost more in the near term than the Department has currently allocated for all of the services' assault aircraft programs. Thus, although the V-22 appears to have advantages of the long term, we know of no way to avoid its greater expense in the short run.

At this time, I would be prepared to present a short presentation on our principal findings, and then Dr. Randall and I would be pleased to answer any questions at all that the committee might have.

Senator INOUYE. Please proceed.

SLIDE 1

ASSESSMENT OF ALTERNATIVES
FOR THE V-22
ASSAULT AIRCRAFT PROGRAM
JUNE 1990

L. DEAN SIMMONS, PROJECT LEADER

SLIDE 2:

CONGRESSIONAL DIRECTION

HOUSE ARMED SERVICES COMMITTEE

THE COMMITTEE DIRECTS THE SECRETARY OF DEFENSE TO PROVIDE WITH THE FISCAL YEAR 1991 BUDGET REQUEST AN INDEPENDENT COST AND OPERATIONAL EFFECTIVENESS ANALYSIS (COEA) OF ALL REASONABLE V-22 ALTERNATIVES INCLUDING, BUT NOT LIMITED TO, THE CH-53E, BV-360, EH-IOI, CH-46E, CH-60 AIRCRAFT OR ANY COMBINATION THEREOF."

Dr. SIMMONS. As you are aware, Mr. Chairman, the original direction for this study was provided by the House Armed Service Committee, which specifically directed the Secretary to provide with the fiscal year 1991 budget request an independent cost-effectiveness assessment of the V-22 and all reasonable alternatives.

SLIDE 3:

CONGRESSIONAL DIRECTION

APPROPRIATIONS CONFERENCE

"… THE COEA STUDY SHOULD ADDRESS ALL ASPECTS OF OPERATIONS THAT COULD BE PERFORMED BY THE V-22 OR ITS ALTERNATIVE, NOT JUST THE OPPOSED AMPHIBIOUS ASSAULT MISSION. THEREFORE, IT IS PARTICULARLY IMPORTANT THAT THE ADVANTAGES OF THE V-22 FOR SPECIAL OPERATIONS FORCES AND DRUG SMUGGUNG INTERDICTION MISSIONS BE EXAMINED IN FULL."

"… INCLUDE THE FOLLOWING KEY FACTORS:

- MISSION ANALYSIS OF MARINE CORPS' AMPHIBIOUS SHIP-TO-SHORE MOVEMENT, LONG-RANGE SPECIAL OPERATIONS, OVER-THE-HORIZON LANDINGS, FOLLOW-ON OPERATIONS, COMBAT SEARCH AND RESCUE OPERATIONS, AND DRUG INTERDICTION OPERATIONS;
- SURVIVABILITY AND THE COST OF ATTRITION;
- COLLATERAL COSTS SUCH AS INCREASED MANPOWER REQUIREMENTS, INCREASED PILOT REQUIREMENT, AND THE IMPACT ON BOTH STRATEGIC AND AMPHIBIOUS LIFT;
- THE DATE WHICH THE V.22 OR AN ALTERNATIVE CAN BE FIELDED AND READY FOR DEPLOYMENT."

In addition, the appropriations conference from the House and the Senate provided some additional direction as to which missions were to be examined and some other factors to be included in the assessment, and this long list of factors accounts, in part, for the considerable length of our study report.

SLIDE 4

OVERVIEW

- AIRCRAFT EXAMINED:

V-22	NEW HELICOPTER (BOEING MODEL 360)
CH-53E+	EH-IOI (UK/ITALY)
CH-46E+	CH-47M
CH-60(S)	SUPER PUMA (FRANCE)

- MISSIONS EXAMINED:

MARINE CORPS	OTHER SERVICE OR AGENCY
AMPHIBIOUS ASSAULT (INCLUDING OVER-THE-HORIZON LANDINGS)	COMBAT SEARCH AND RESCUE (NAVY)
SUBSEQUENT OPERATIONS ASHORE (INCLUDING RESUPPLY TO FORWARD DEPLOYED FORCES)	LONG-RANGE SPECIAL OPERATIONS (AIR FORCE)
DEPLOYMENT MISSIONS	DRUG INTERDICTION
HOSTAGE RESCUE OR RAID	ANTISUBMARINE WARFARE (NAVY)

- COST-EFFECTIVENESS ASSESSED BY CAPABILITIES OF EQUAL-COST AIRCRAFT FLEETS AT TWO FUNDING LEVELS

A quick overview: we examined eight different aircraft, singly and in combination, to provide the various assault aircraft missions—the V-22, all of the helicopter alternatives cited in the House Armed Services Committee directive, and two alternatives that we thought reasonable based on their capabilities (the Army's CH-47 Chinook, a marinized version of that aircraft, and the French firm Aerospatiale's Super Puma aircraft, which has about the same characteristics as some of the other helicopters we were examining). We examined eight different missions for the aircraft—four for the Marine Corps, four for other service or government agency missions. All of those had been identified in the directions provided by the Congress to the Secretary.

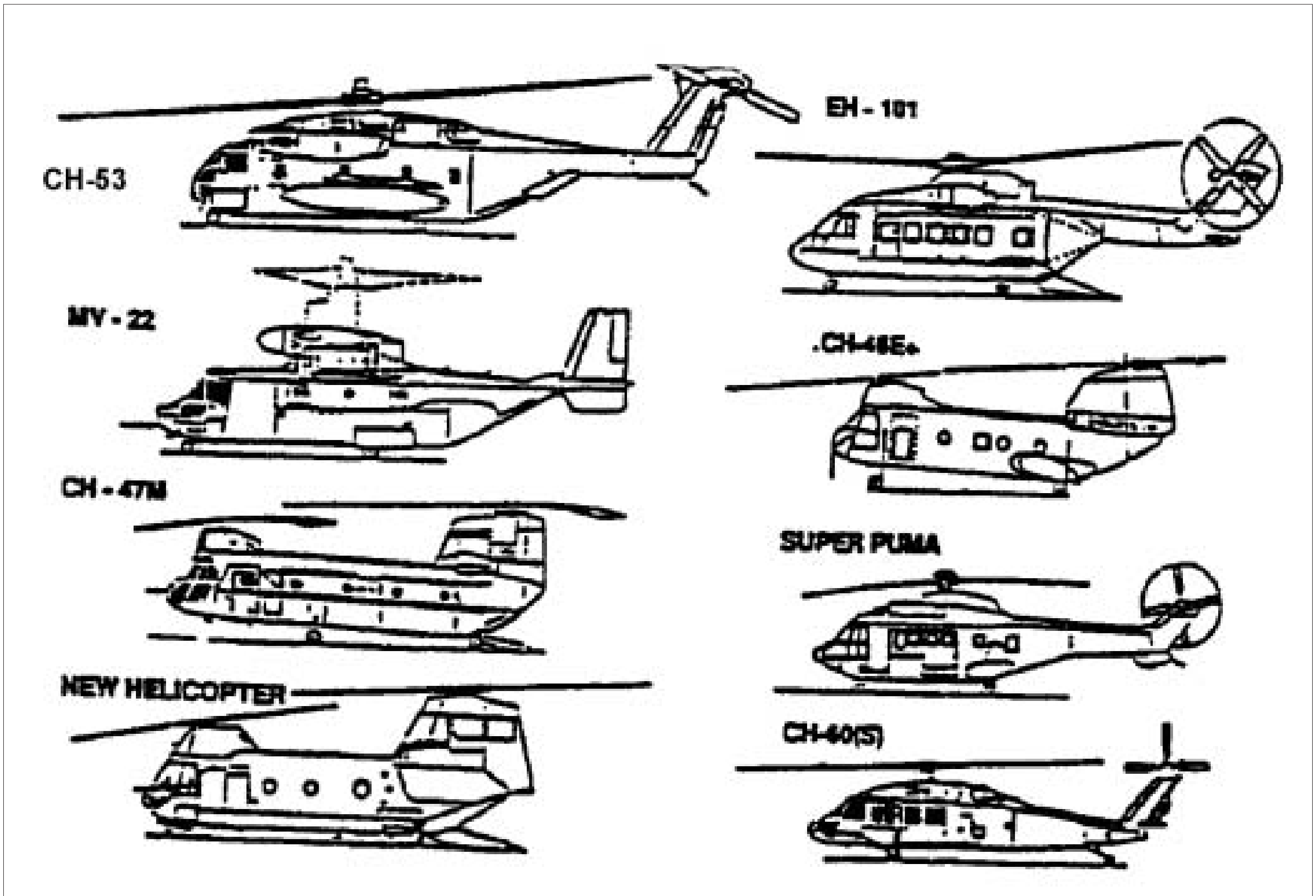
As I mentioned in my opening statement, we compared the cost-effectiveness of the alternatives by looking at the capabilities of aircraft fleet sizes with equal 20-year system cost.

SLIDE 5:

The outline for the remainder of the presentation is shown here. I will discuss briefly the aircraft, spend a considerable amount of time on the Marine Corps missions, discuss the four other service missions, discuss briefly the near-term costs, and summarize with two quick overview charts.

ALTERNATIVE AIRCRAFT

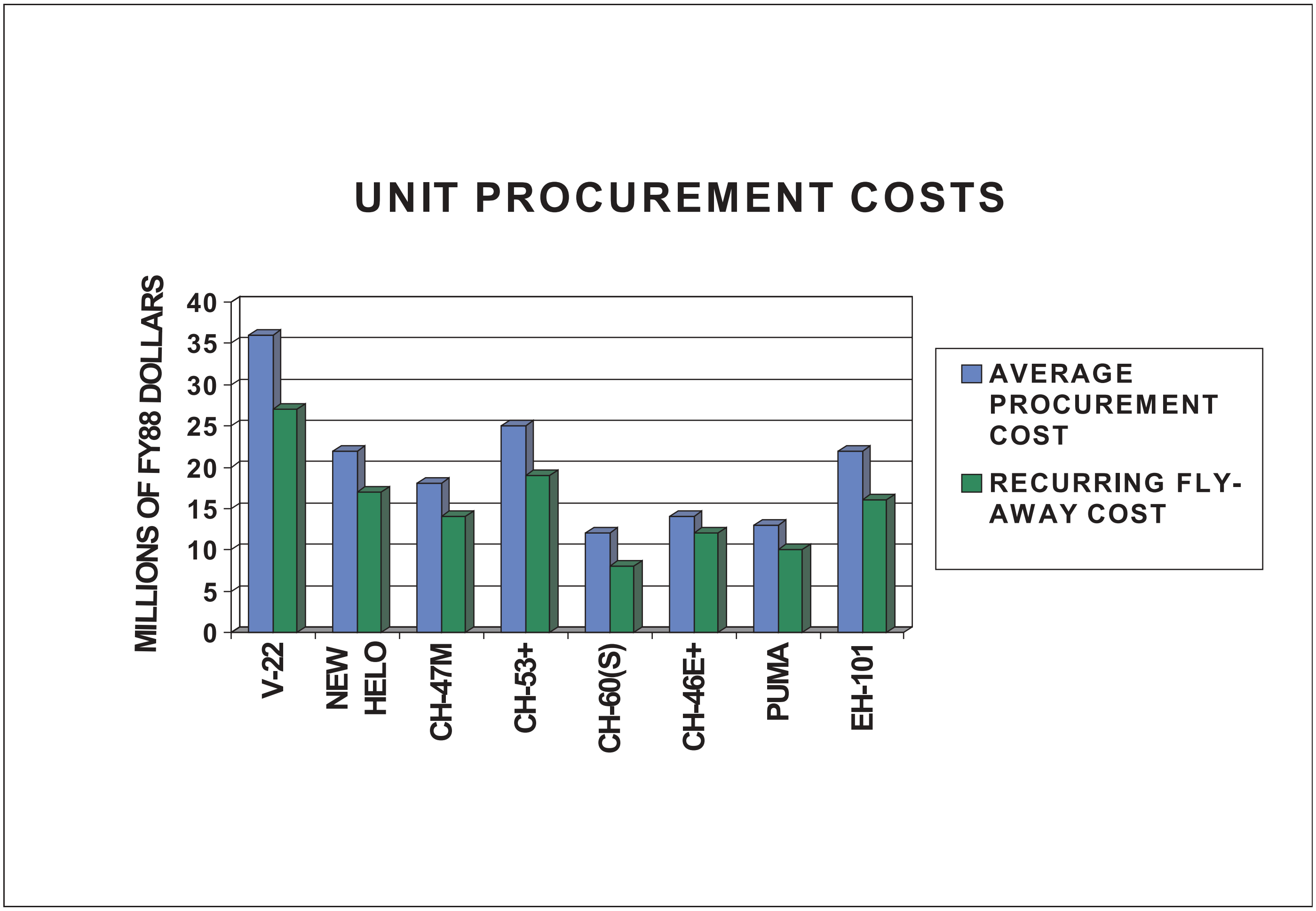
SLIDE 6:



Here, shown to the same scale, are all of the alternative aircraft that we examined. The V-22 is the second aircraft on the left side of the chart. It is the only non-helicopter aircraft that we considered, being a tilt-rotor aircraft. We show the standard airplane configuration of the rotor and the helicopter configuration of the rotor.

The largest aircraft we considered is the Marine Corps' current CH-53E heavy-lift helicopter. The Army's Chinook is shown here, a new helicopter modeled on Boeing's 360 technology demonstrator, the British-Italian EH-101, the Marines' current CH-46, the Aerospatiale Super Puma, and the Army's CH-60 Blackhawk helicopter.

SLIDE 7



The difference in costs for the aircraft are illustrated here, where we show the unit procurement cost, both the average procurement cost and the fly-away cost for the aircraft. The costs for the V-22 are estimated to be substantially larger than those for the other aircraft, an average of about $35 million per V-22 over the entire program buy that has been proposed. The next most expensive aircraft would be the CH-53 heavy-lift helicopter, at about $25 million. As you can see, some of the helicopters have cost only on the order of $l0 million each.

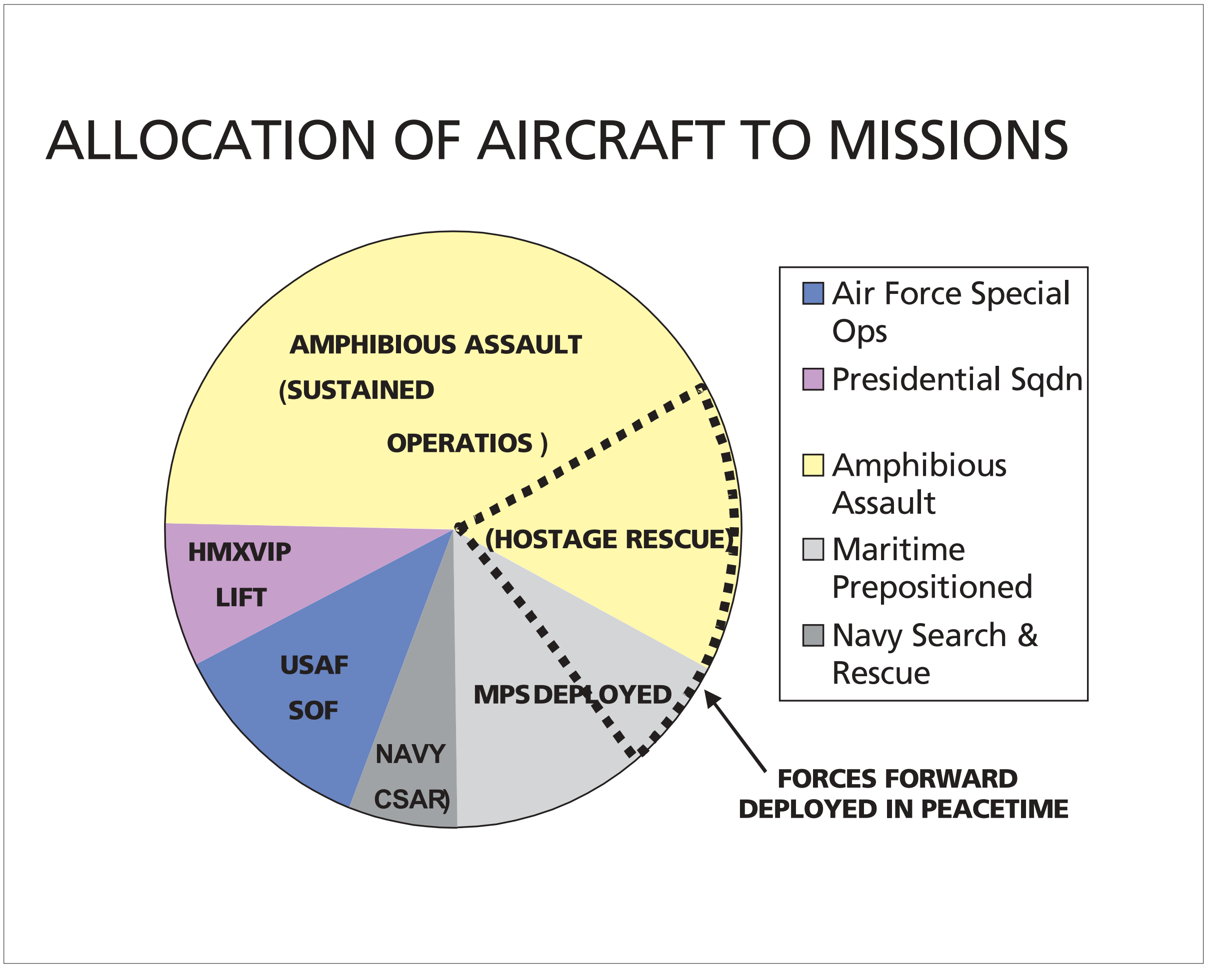
SLIDE 8:

MARINE CORPS MISSION

MISSION	CONTRIBUTION OF ASSAULT AIRCRAFT
AMPHIBIOUS ASSAULT	MOVE TROOPS AND EQUIPMENT ASHORE
SUSTAINED OPERATIONS	MOVE TROOPS AND EQUIPMENT TO SUPPORT COMBAT FORCES ASHORE
HOSTAGE RESCUE OR RAID	INSERT AND EXTRACT MARINE RESCUE OR RAIDING FORCE OR HOSTAGES
OVERSEAS DEPLOYMENT (NALMEB OR MPFMEB)	MOVE TO OVERSEAS THEATER AND TRANSPORT DEPLOYED MARINE FORCE TO COMBAT POSITIONS

At this point, I would like to move to a discussion of the cost effectiveness for the Marine Corps missions.

SLIDE 9:

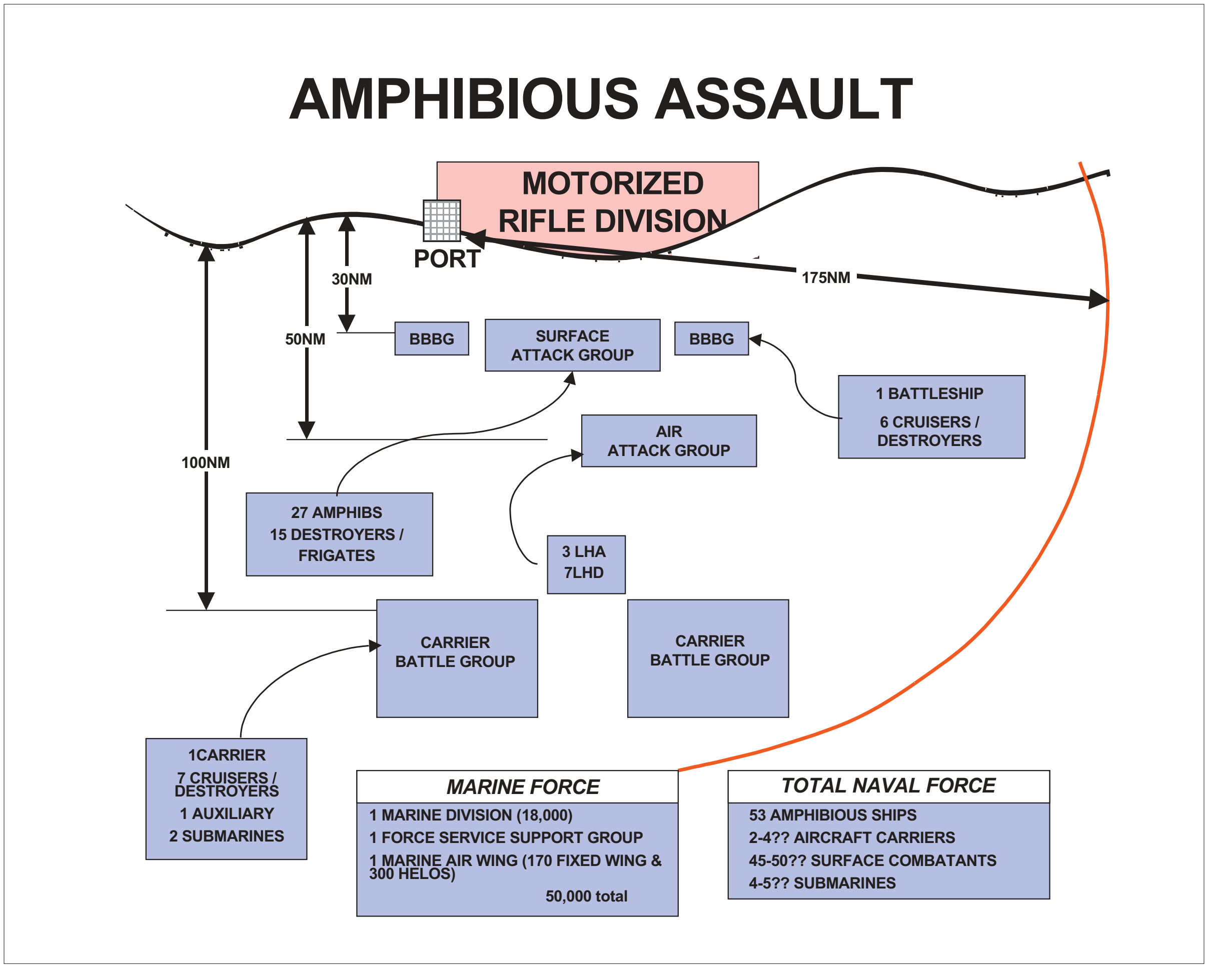


To give you an idea of how the aircraft are supposed to be used, on this chart we show the percent distribution of aircraft among the potential missions. Of the planned buy, about 60 percent of the aircraft would be designated for the amphibious assault or the sustained operations ashore mission. This is shown in the top portion of the chart that extends between the two solid black lines. An additional one-fourth or so of the total aircraft would be designated for the Marines' maritime prepositioned forces. For deployment missions they have two brigades, plus a brigade that is designated for potential use on NATO's northern flank in that context. In addition, we show the peacetime forward-deployed forces that we modeled as carrying out a hostage rescue or small-scale combat raid mission.

The Marine Corps had also planned to buy 20 aircraft for the Presidential support squadron to replace the existing aircraft that are used there. Two other services, the Navy and the Air Force, also proposed to use the V-22—the Navy for combat search and rescue and the Air Force for special operations.

In our assessment, we measured capabilities for all of these potential uses for the aircraft with the exception of the Presidential support squadron.

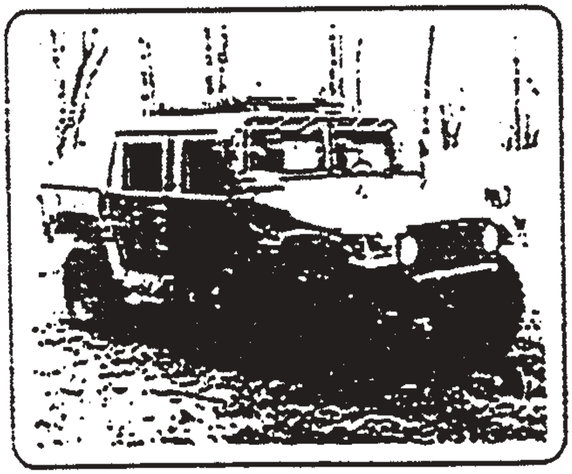
SLIDE 10:



The first mission we looked at is the amphibious assault. The scenario we used envisions a division-size amphibious assault supported by the types of military forces shown here; carrier battle forces from the Navy, the amphibious assault ships shown in the tan or light brown blocks, and other Navy supporting forces that would provide protection for the amphibious forces. The threat force is postulated as a division-size force deployed over the red area shown. In carrying out this assault, the Marine forces would be landed in two ways: one-half by surface assault vehicles and landing craft, and the other one-half of the assault force by assault aircraft the V-22 or one of the helicopter alternatives.

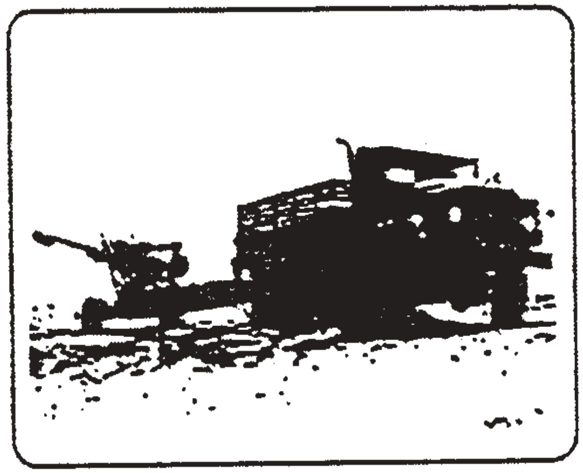
SLIDE 11

HIGH MOBILITY MULTIPURPOSE WHEELED VEHICLE (HMMWV)



WEIGHT: 7,500 - 9,150 LB

M923 TRUCK AND Mi98 HOWITZER



TRUCK WEIGHT: 21,800 LB
HOWITZER WEIGHT: 18,250 LB

The next chart will give you an idea of the kind of equipment that is included in the Marines' vertical assault force.

The largest number of items of equipment are [sic.] the high-mobility, multipurpose, wheeled vehicles, or HMMWV's. The typical Marine vertical assault force includes about 300 of these vehicles, which are too large to fit inside any of the aircraft and must be carried in sling loads externally.

The other major vehicles for the airlifted force are the trucks and the artillery pieces that they tow to provide fire support for the Marine forces. Those two pieces of equipment are so heavy that only the large CH-53E helicopter can carry them.

SLIDE 12:

AIRCRAFT FLEET ALTERNATIVES

HEAVY-LIFT AIRCRAFT

- ALL FLEETS INCLUDE PLANNED CH-53E FORCE

OTHER AIRCRAFT CAPABLE OF CARRYING HMMWVs

- V-22
- NEW HELICOPTER
- CH-47M

AIRCRAFT NOT CAPABLE OF CARRYING HMMWVs - ADDITIONAL CH-53Es REQUIRED

- CH-60 (STRETCHED)
- CH-46E+
- SUPER PUMA
- EH-101

So, based on this equipment, we have constructed fleets of aircraft that we then compare in our effectiveness assessments. Because we had those large items of equipment, we assumed that all of our fleets would include the Marine Corps' planned CH-53E large helicopter force.

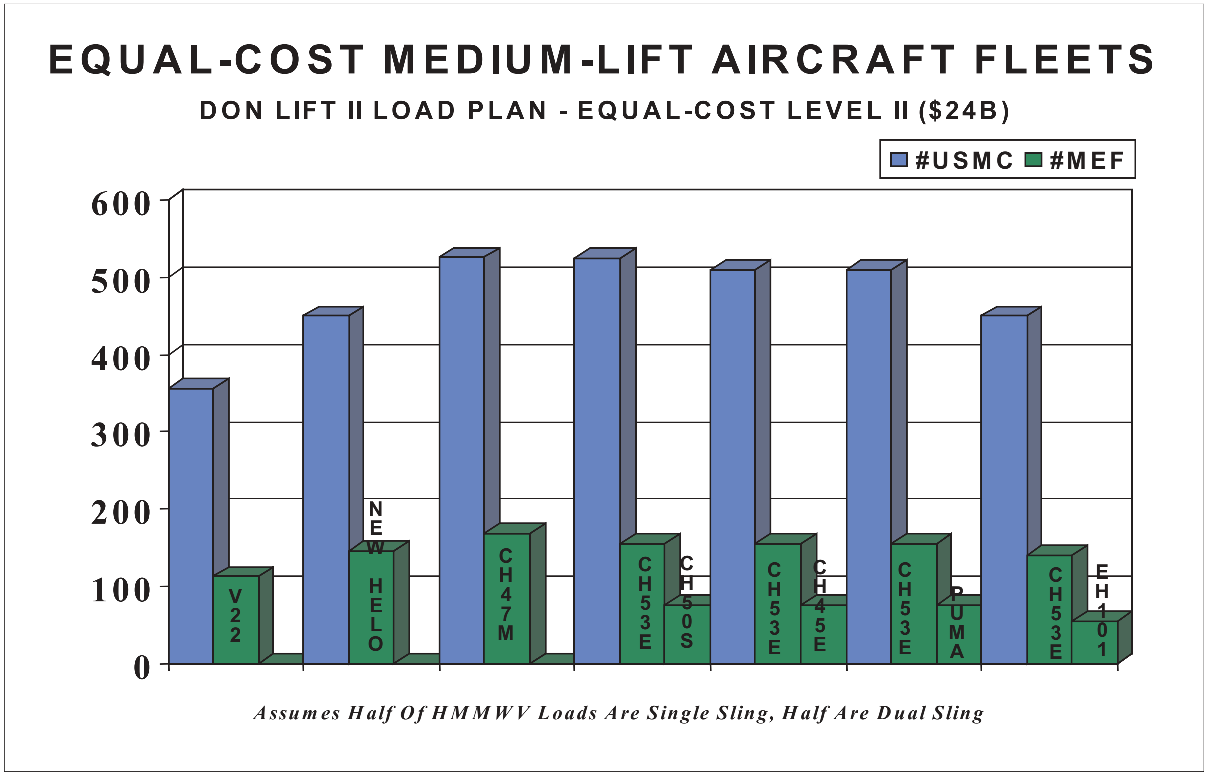
The other aircraft fall into two groups. [First,] those that are large enough to carry the HMMWV as an external load. The V-22, a new design helicopter, and the CH-47 would fall into that class. Four of the other aircraft are too small to carry the HMMWV and in order to use those aircraft another helicopter would have to be provided in the mix. The most efficient way of doing that would be to add additional CH-53E's, so that is the assumption we made.

The other aircraft are the CH-60, the CH-46, the Super Puma, and the EH-101. When we discuss alternatives, we will identify some of them by a single aircraft name—those are the ones that can carry the HMMWV. Other alternatives are designated by two names. That means that two types of aircraft are required, one to carry the Marine combat troops and one to carry the heavier equipment.

At this time, I would like to show you how we took the costs and constructed equal cost fleets.

(There is no SLIDE 13 in the Congressional Record.)

SLIDE 14:



The cost level that we have chosen to show you is what we call cost level II in our study, or $24 billion, the approximate cost of a fleet of CH-53E and CH-60 helicopters that would be comparable to the fleet that the Department of Defense substituted for the V-22 when that program was canceled.

Given the differences in the cost of the aircraft, you see on this chart that the number of V-22's that we can buy is lower than the numbers for any of the other alternatives. For this $24 billion, we could buy 356 V-22's and support them over 20 years of operation.

The lower portion of the bar designates the number of aircraft that would be available for the Marine Expeditionary Force, the division-size force that we assumed for our effectiveness analysis.

[There is no SLIDE 15 in the Congressional Record.]

MARINE CORPS MISSIONS COST EFFECTIVENESS

Now I would like to discuss the effectiveness or the combat capabilities of those numbers of aircraft.

SLIDE 16:

MARINE CORPS MISSION	
MISSION	**CONTRIBUTION OF ASSAULT AIRCRAFT**
AMPHIBIOUS ASSAULT	**MOVE TROOPS AND EQUIPMENT ASHORE**
SUSTAINED OPERATIONS	**MOVE TROOPS AND EQUIPMENT TO SUPPORT COMBAT FORCES ASHORE**
HOSTAGE RESCUE OR RAID	**INSERT AND EXTRACT MARINE RESCUE OR RAIDING FORCE OR HOSTAGES**
OVERSEAS DEPLOYMENT (NALMEB OR MPFMEB)	**MOVE TO OVERSEAS THEATER AND TRANSPORT DEPLOYED MARINE FORCE TO COMBAT POSITIONS**

First, for the amphibious assault mission in which the aircraft are used to move marines and their combat equipment ashore.

SLIDE 17

FACTORS EXAMINED IN COMBAT EFFECTIVENESS ASSESSMENT

FORCE STRUCTURE AND COMPOSITION FACTORS

- NUMBER AND MIX OF AIRCRAFT
- COMPOSITION AND SIZE OF MARINE ASSAULT FORCE
- TYPE AVIONICS USED IN ASSAULT AIRCRAFT
- TYPE VEHICLE USED AS HMMWV

TACTICAL FACTORS

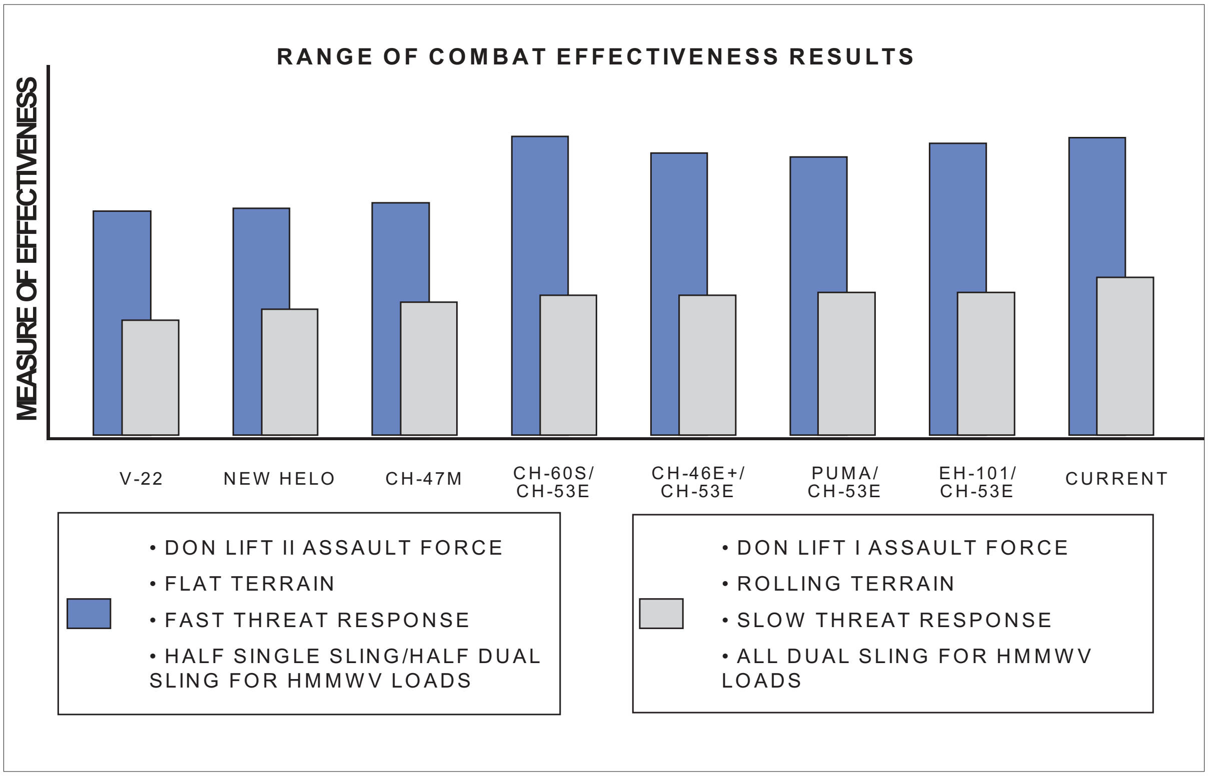
- STANDOFF DISTANCE
- HMMWV CARRIAGE
- TROOPS CARRIED IN CH-53Es AND CH-47s
- TIME OF ASSAULT

TERRAIN AND THREAT FACTORS

- TYPE TERRAIN
- THREAT FORCE SIZE
- THREAT FORCE RESPONSE
- NO AIRCRAFT ATTRITION

In this portion of the assessment, we examined a broad range of factors that could potentially influence the results. The wide set of variations that we considered is another of the reasons why our report is so large. We looked at differences in force structure and composition, the number of aircraft in the fleet, the composition of the Marine assault force itself, the type of avionics used in the aircraft, and the type of vehicle used as a HMMWV. We also examined a wide range of tactical factors, and similarly a wide range of potential operating terrain and threat factors.

SLIDE 18:



The next chart gives you an idea of the range of combat effectiveness results that we obtained. The measure of effectiveness that we used here was the fraction of the marine assault force lost during the combat operation, both lost in assault aircraft en route to their landing zones and subsequent losses during combat operations ashore. So in this case, a lower number represents a more effective option.

We show two sets of results to give you an idea of the range of values that we obtained in our assessment. Under the first set of conditions, identified by the lower bars, there is very little difference among the alternatives, although if you look closely you will see that the losses incurred by the V-22 force are somewhat lower than those for the other alternatives. For a different set of conditions, particularly when the survivability features of the V-22 are more important, then the differences in losses are considerably larger.

SUSTAINED OPERATIONS

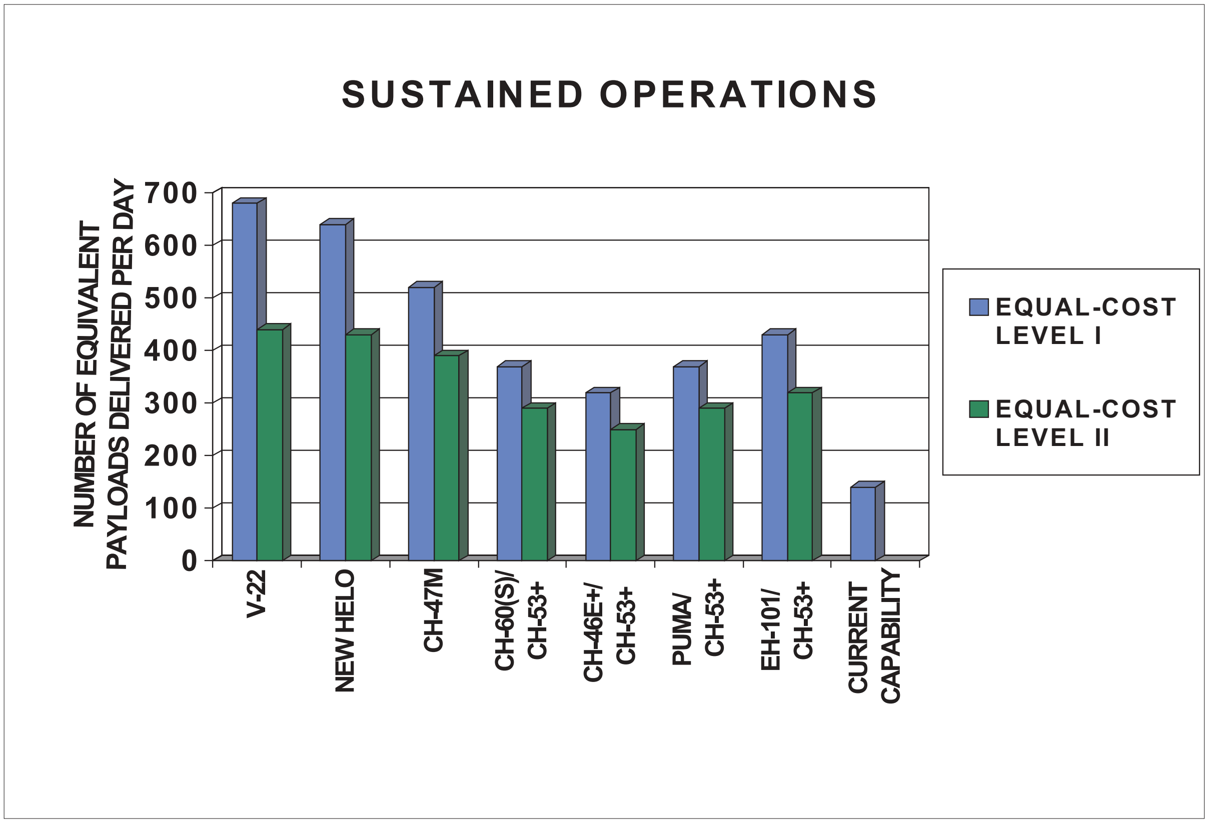
SLIDE 19:

MARINE CORPS MISSION

MISSION	CONTRIBUTION OF ASSAULT AIRCRAFT
AMPHIBIOUS ASSAULT	MOVE TROOPS AND EQUIPMENT ASHORE
SUSTAINED OPERATIONS	MOVE TROOPS AND EQUIPMENT TO SUPPORT COMBAT FORCES ASHORE
HOSTAGE RESCUE OR RAID	INSERT AND EXTRACT MARINE RESCUE OR RAIDING FORCE OR HOSTAGES
OVERSEAS DEPLOYMENT (NALMEB OR MPFMEB)	MOVE TO OVERSEAS THEATER AND TRANSPORT DEPLOYED MARINE FORCE TO COMBAT POSITIONS

The next mission that we examined was the sustained operations mission. During this mission it is assumed that the aircraft would have transitioned to shore from the amphibious lift ships where they would have been used for the amphibious mission, and at this time they would be used to support combat operations ashore.

SLIDE 20:



The measure of effectiveness we used here [slide 20] was the number of equivalent payloads that could be delivered per day over a 30-day period. Results are shown for both of the cost levels that we examined in the assessment—we have been focusing on the results obtained at cost level II at this point. Those are shown in the lower set of bars. Again, the V-22 provides more capability to deliver equivalent size payloads over the 30-day period that we examined.

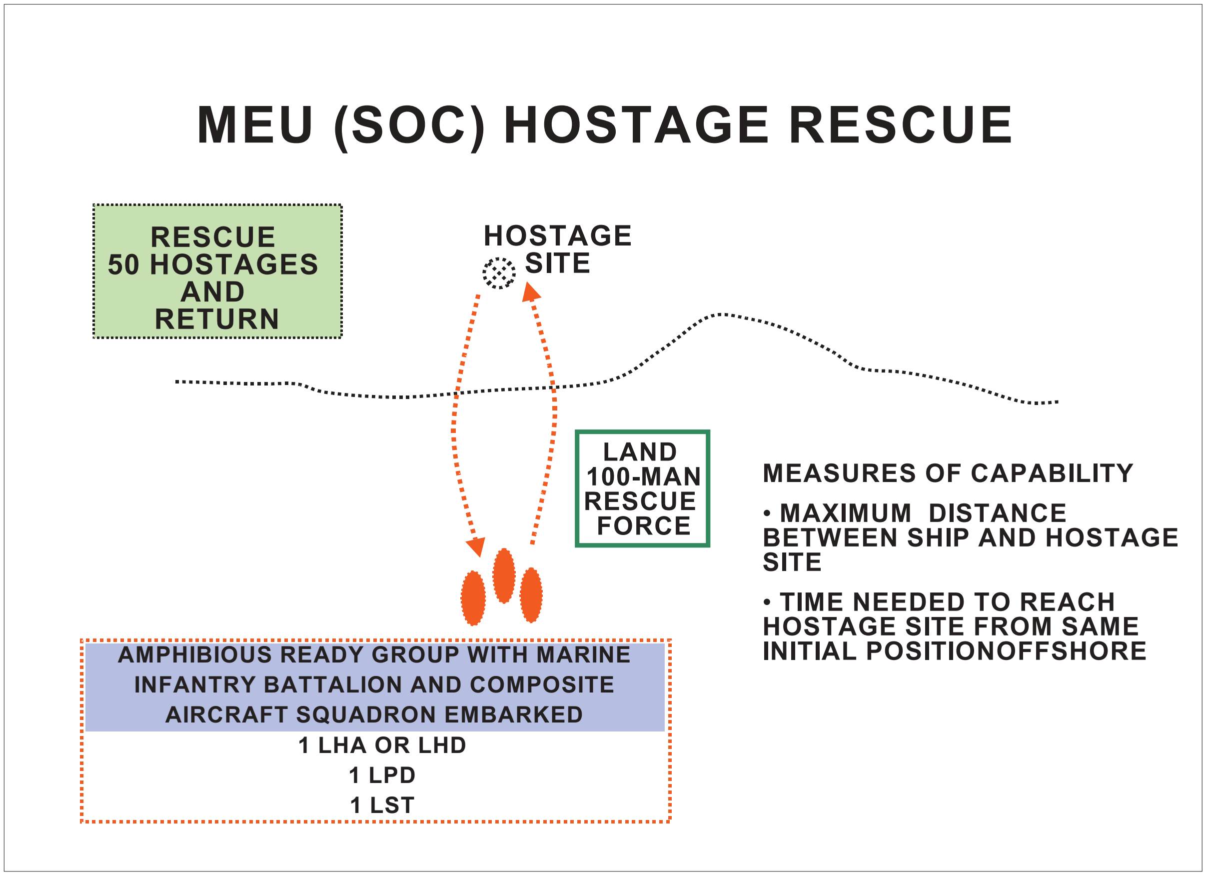
You will also note, though, that all of the alternatives provide a considerable improvement over the Marines' current capability.

SLIDE 21:

MARINE CORPS MISSION	
MISSION	**CONTRIBUTION OF ASSAULT AIRCRAFT**
AMPHIBIOUS ASSAULT	MOVE TROOPS AND EQUIPMENT ASHORE
SUSTAINED OPERATIONS	MOVE TROOPS AND EQUIPMENT TO SUPPORT COMBAT FORCES ASHORE
HOSTAGE RESCUE OR RAID	INSERT AND EXTRACT MARINE RESCUE OR RAIDING FORCE OR HOSTAGES
OVERSEAS DEPLOYMENT (NALMEB OR MPFMEB)	MOVE TO OVERSEAS THEATER AND TRANSPORT DEPLOYED MARINE FORCE TO COMBAT POSITIONS

The next mission that we examined was the small-scale hostage rescue or combat raid. In this mission the aircraft would be used to insert and extract a Marine rescue or a raiding force, and in the case of the rescue, to also extract hostages.

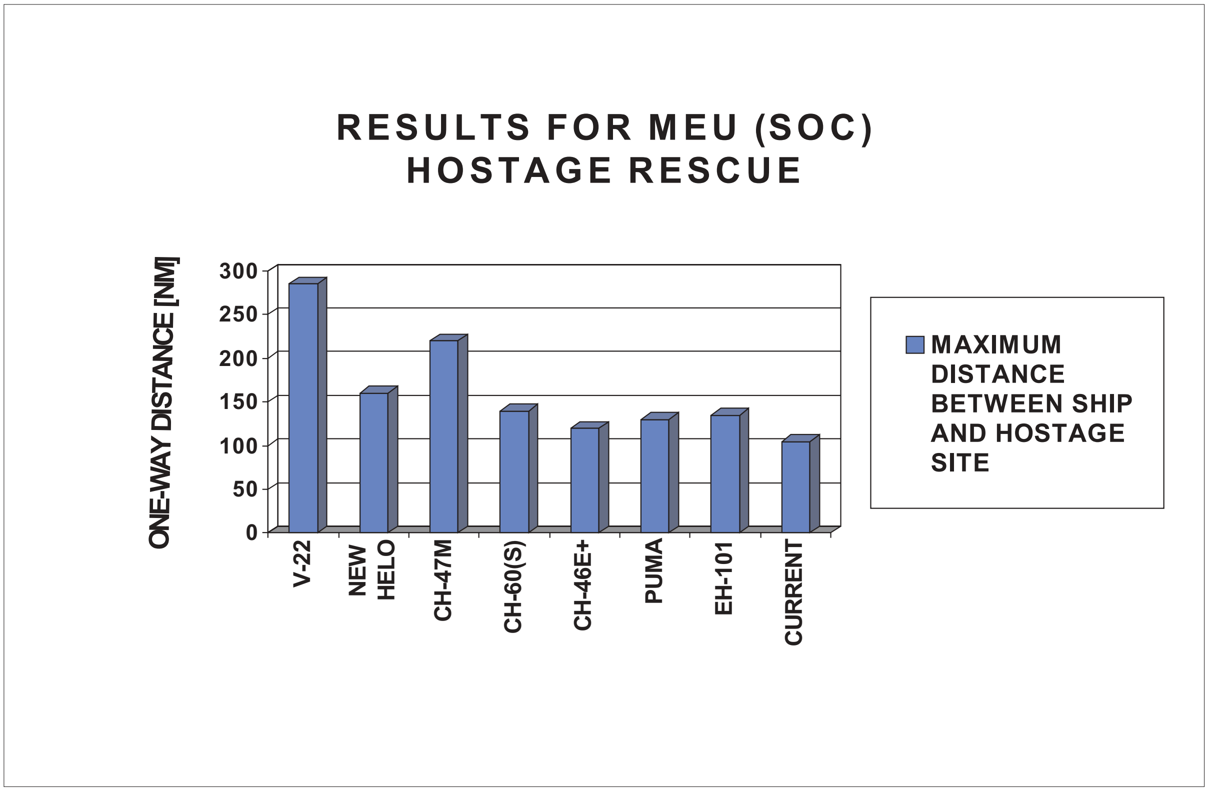
SLIDE 22:

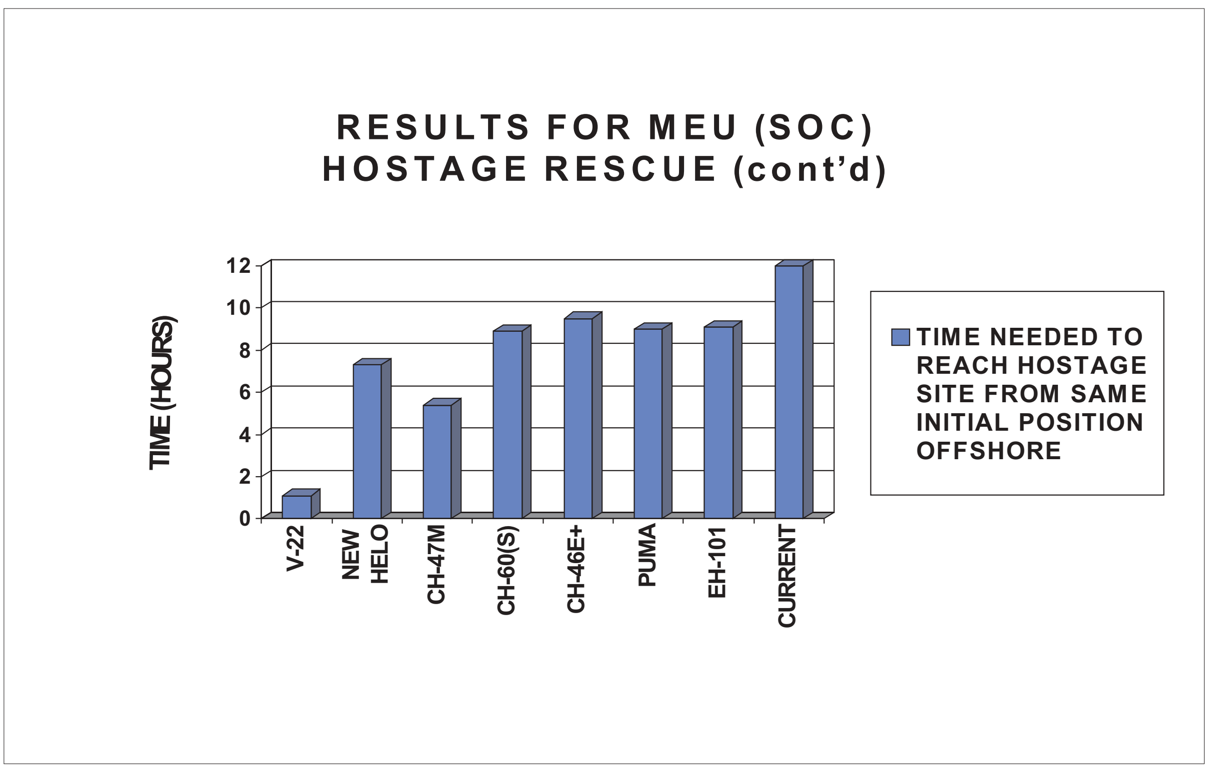


How this operation would be conducted is depicted on this chart. It is assumed that it would be carried out by the forces deployed with a Marine amphibious ready group. We have two such units forward deployed at all times. During peacetime conditions it consists of an infantry battalion and a composite aircraft squadron that would include the assault aircraft and typically three ships. We assumed that the aircraft would be launched from these ships into some hostage site or a location for a combat raid. The rescue or raiding force would consist of 100 marines and they would rescue some number of hostages.

We measured the capability for this mission by looking at the maximum distance that we could be away from the raiding site and conduct the operation and the time it would take to reach that site from the same starting position offshore.

SLIDE 23:





The results are depicted on this chart. The maximum distance between the ship and the hostage site is shown in the left-hand set of bars. Distance is measured in nautical miles and the various alternatives are listed along the bottom of the slide. The speed and range of the V-22 would allow it to conduct this operation from substantially greater distances off-shore or to substantially greater distances on land than would be the case with any of the other alternatives.

Given this advantage in stand-off distance, if we were to start the operation from the same point and we chose the largest of those distances as our starting point, the V-22 would have a five to perhaps 12-hour advantage in the amount of time it would take to reach the hostage site from the same starting position offshore.

The last Marine Corps mission we examined was the overseas deployment that would be

SLIDE 24:

MARINE CORPS MISSION

MISSION	CONTRIBUTION OF ASSAULT AIRCRAFT
AMPHIBIOUS ASSAULT	MOVE TROOPS AND EQUIPMENT ASHORE
SUSTAINED OPERATIONS	MOVE TROOPS AND EQUIPMENT TO SUPPORT COMBAT FORCES ASHORE
HOSTAGE RESCUE OR RAID	INSERT AND EXTRACT MARINE RESCUE OR RAIDING FORCE OR HOSTAGES
OVERSEAS DEPLOYMENT (NALMEB OR MPFMEB)	MOVE TO OVERSEAS THEATER AND TRANSPORT DEPLOYED MARINE FORCE TO COMBAT POSITIONS

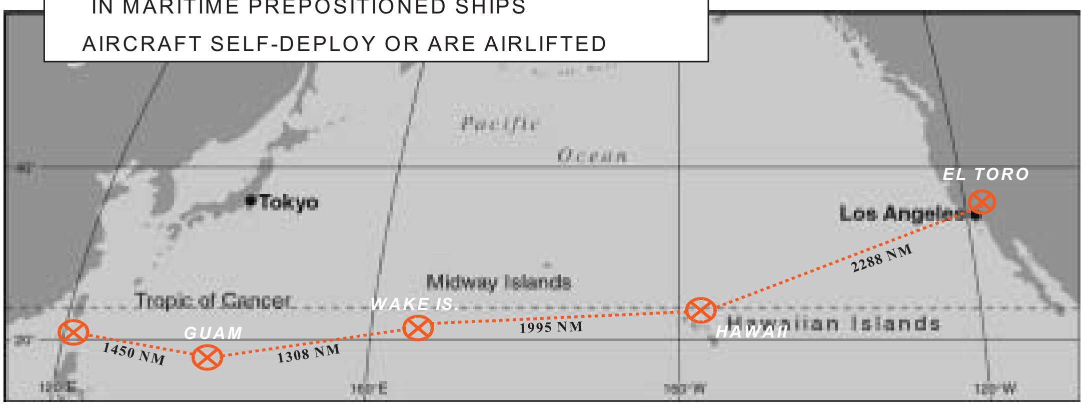
used with our Norway air-landed brigade to move a brigade-size force to NATO's northern flanks or with one of our maritime prepositioned force brigades to operate in conjunction with the Navy's maritime prepositioned squadrons.

In this case, the assault aircraft would be moved to the overseas theater either by self-deploying in the case of the V-22 or being airlifted in strategic airlift aircraft. Once overseas, the aircraft would deploy Marine forces to their combat positions.

SLIDE 25:

DEPLOYMENT ROUTE FOR MPFMEB AIR COMBAT ELEMENT

MARINE BRIGADE AIRLIFTED
EQUIPMENT AND SUPPLIES FORWARD DEPLOYED IN MARITIME PREPOSITIONED SHIPS
AIRCRAFT SELF-DEPLOY OR ARE AIRLIFTED

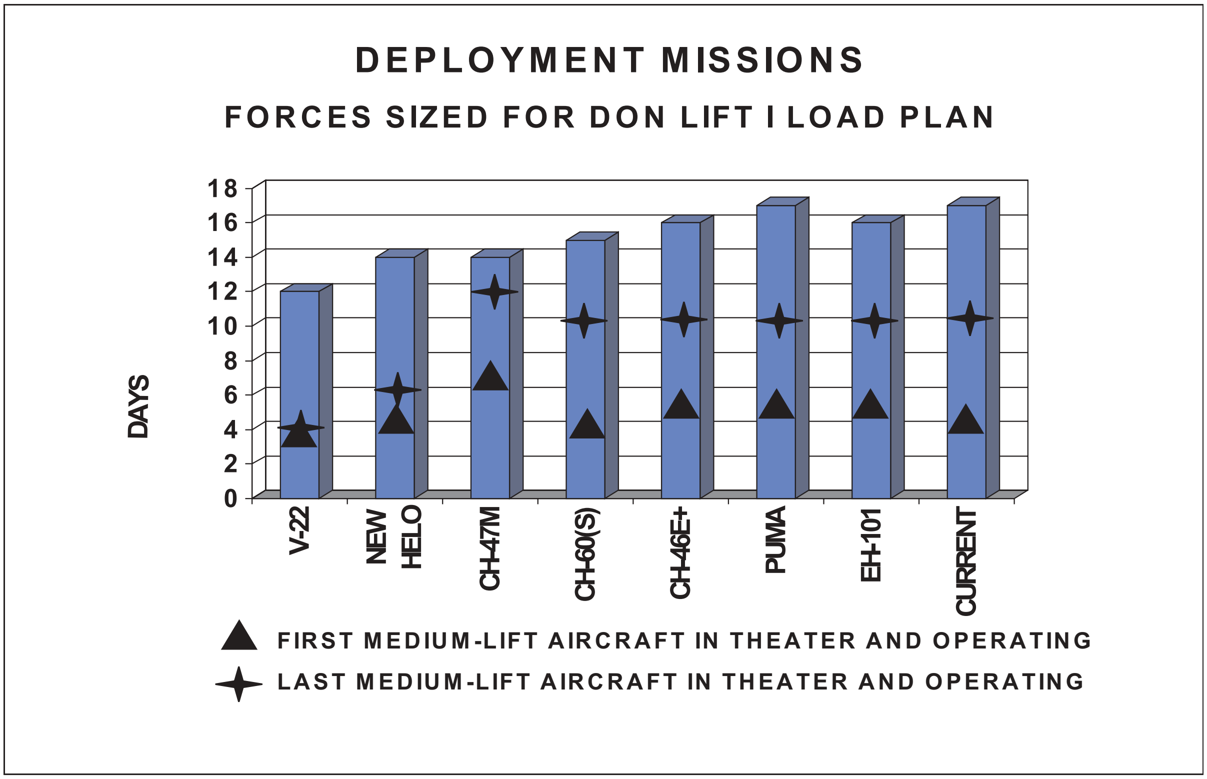


The typical deployment scenario is shown here. Moving from a Marine Corps base at El Toro, CA, to a combat site in the Philippines with intermediate stops as indicated in Hawaii, Wake, and Guam, the Marine infantry forces would be airlifted by Air Force aircraft. The equipment and supplies are already forward deployed in maritime prepositioning ships that are based at Guam in peacetime. The aircraft would self-deploy or be airlifted.

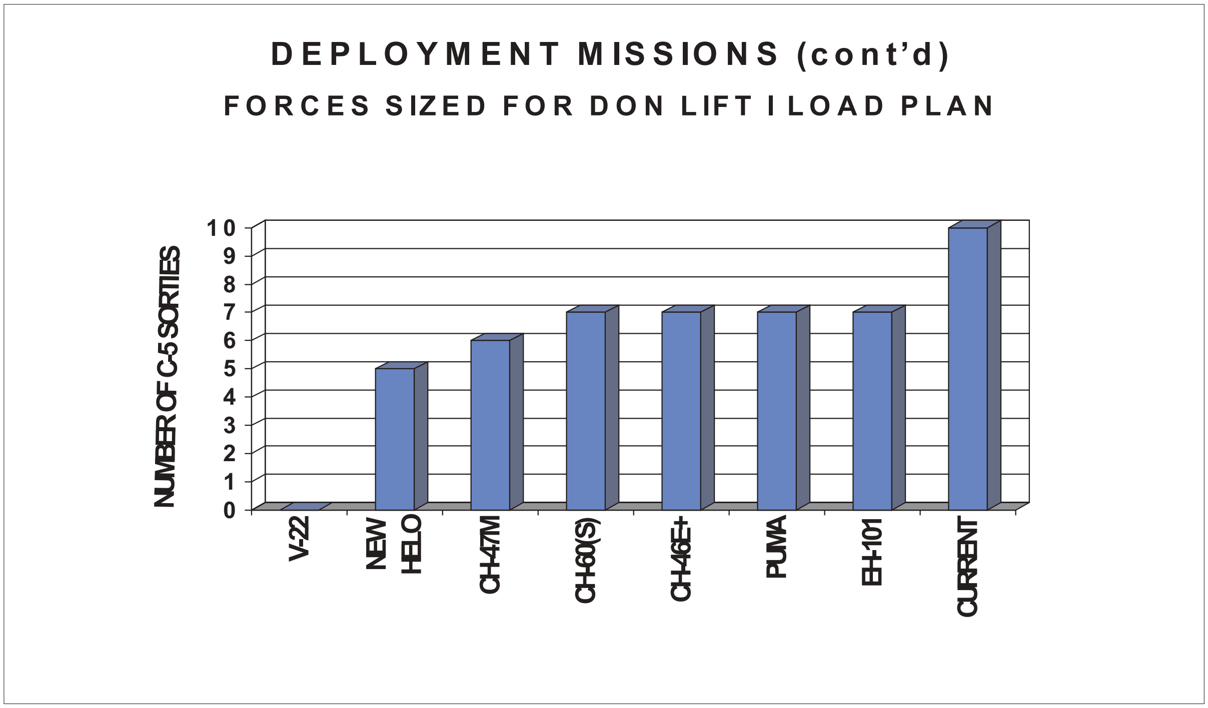
The results for that mission are shown here.

The time to complete the aircraft deployment and move troops is shown in the left-hand set of bars, the faster time giving you an advantage. The V-22, because it can self-deploy and the helicopter alternatives cannot, is able to reach the theater sooner, as indicated by the location of the two triangles, and complete the deployment of troops in the theater more rapidly than any of the helicopter alternatives.

SLIDE 26:



SLIDE 26 CONTINUED



The helicopters are at a disadvantage in this mission because they do not have sufficient range to self-deploy. At these distances, they would have to be disassembled, loaded into strategic airlift aircraft, flown to the theater, and then reassembled and flight tested be-fore they could be used. That difference is indicated by the number of C-5 sorties required. The V-22 would require no sorties because it can self-deploy.

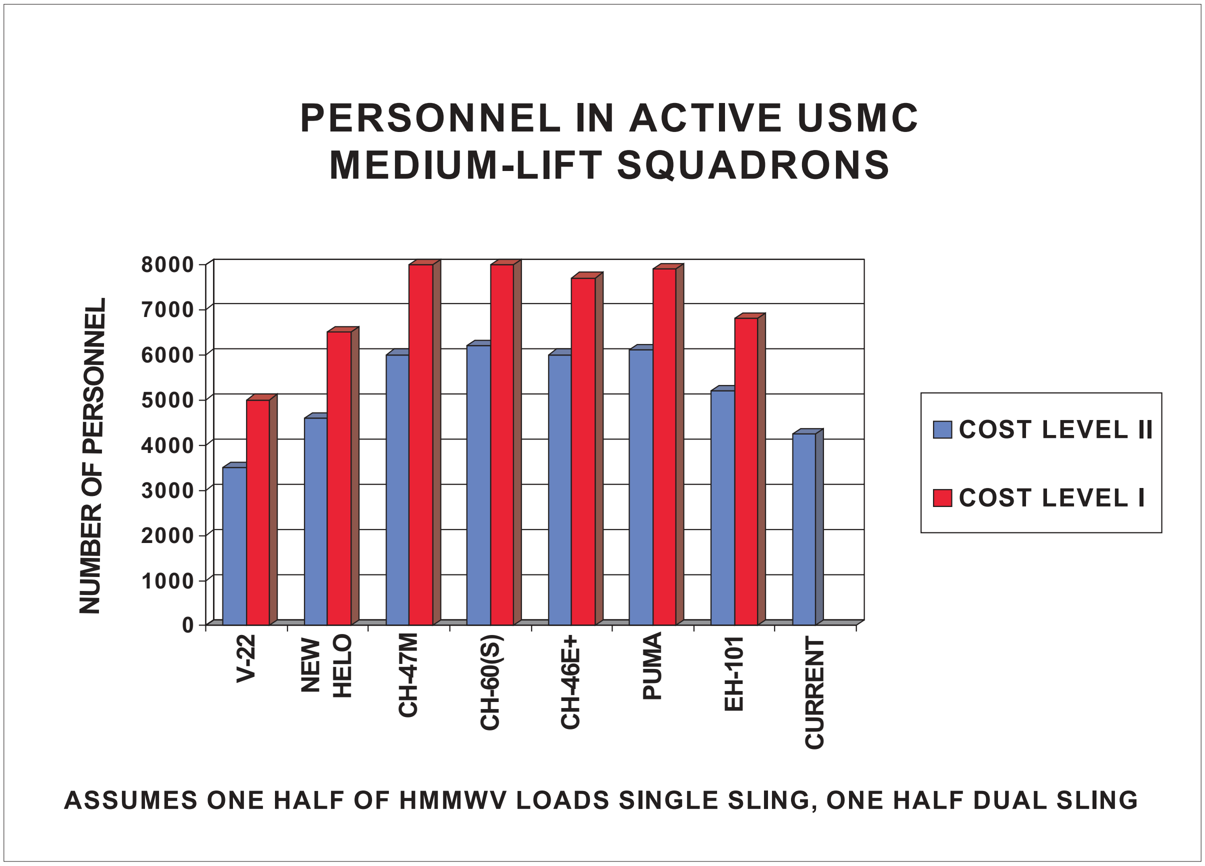
Moving any of the helicopter alternatives would require 5 to 10 sorties to move the air assault portion of the brigade force.

[There is no SLIDE 27 in the Congressional Record.]

MARINE CORPS MISSIONS
OTHER FACTORS

One other important factor with regard to the Marine Corps mission that we wanted to discuss is the manpower requirements.

SLIDE 28:



We show them on this chart. This is the number of personnel in the Marine Corps' active medium-lift helicopter squadrons. The number in those units today is shown by the bar on the right; about 4,500 Marines are assigned to those units. At the higher of the two cost levels examined all of the alternatives would add additional Marines to those units, with the V-22 adding the least about 500 additional personnel.

At cost level II, the $24 billion level we have been focusing on in our briefing, the V-22 would allow a reduction of about 700 Marines from current levels. All of the other alternatives for an equal cost force would require the addition of more Marines for this mission.

OTHER SERVICE OR AGENCY MISSIONS

Now I would like to quickly discuss the other service or agency missions that we examined. [Slide 29 omitted from the record.]

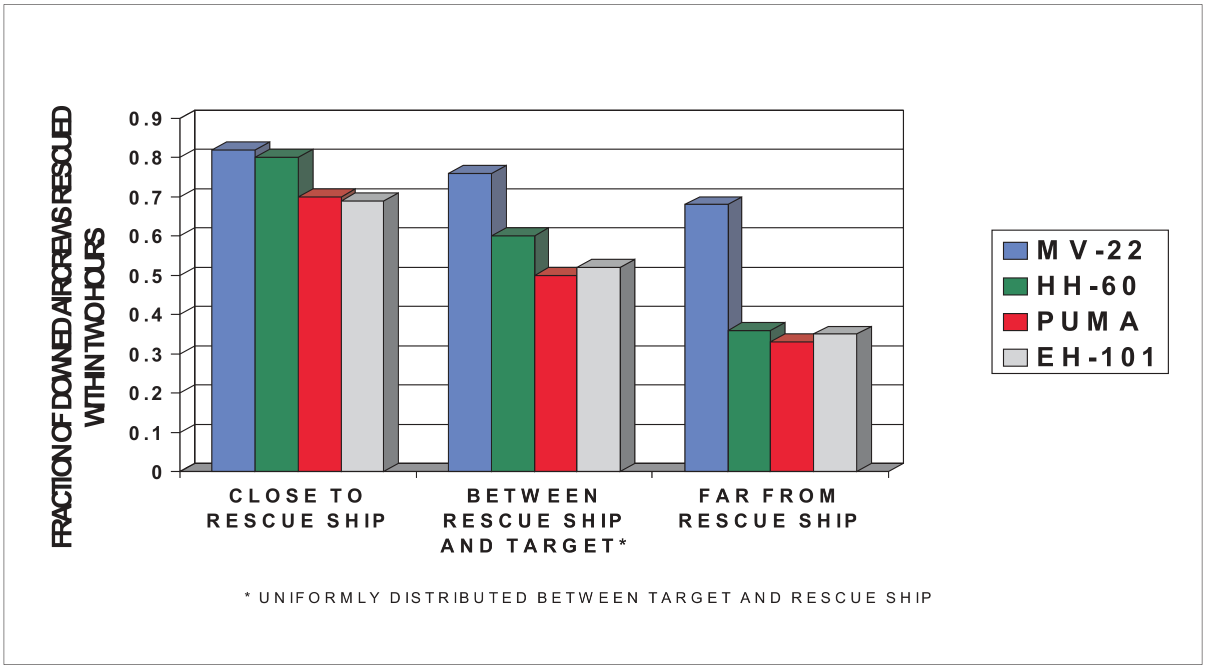
[There is no SLIDE 29 in the Congressional Record.]

SLIDE 30:

OTHER SERVICE MISSIONS	
MISSION	CONTRIBUTION OF ASSAULT AIRCRAFT
COMBAT SEARCH AND RESCUE	RECOVER DOWNED AIRCREWS
SPECIAL OPERATIONS	INSERT AND EXTRACT SPECIAL OPERATIONS FORCES
DRUG INTERDICTION	TRAIL COURIER AIRCRAFT AND BOATS, DEPLOY LAW ENFORCEMENT PERSONNEL
ANTISUBMARINE WARFARE	DETECT AND ATTACK SUBMARINES

The first of these is the Navy's combat search and rescue mission. The aircraft are used to recover downed aircrews from strike or fighter aircraft.

SLIDE 31:



The range of results we obtained for that mission are indicated on this chart. The measure that we used to compare capabilities is the fraction of the downed crews that can be rescued by the different aircraft within two hours. We looked at four alternatives: the V-22, the H-60, the Puma, and the EH-101.

Again, the range of results shown are for different conditions. If all of the aircrews are downed close to the rescue ship, the V-22 has only a slight advantage over the H-60. But as the distances increase, the speed of the V-22 gives it an additional capability relative to the helicopters. In the most extreme conditions that we examined, the V-22 would have a factor of two advantage in the fraction of crews that it could rescue.

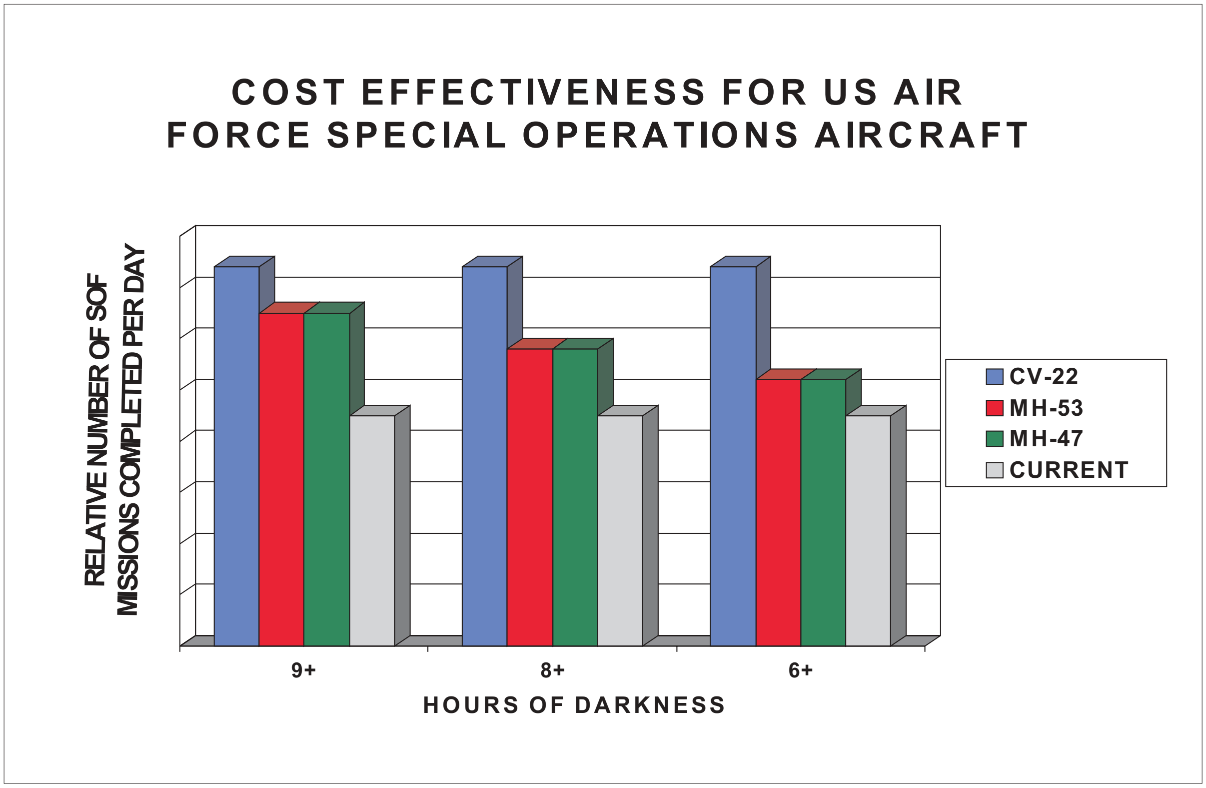
AIR FORCE SPECIAL OPERATIONS

SLIDE 32:

OTHER SERVICE MISSIONS	
MISSION	CONTRIBUTION OF ASSAULT AIRCRAFT
COMBAT SEARCH AND RESCUE	RECOVER DOWNED AIRCREWS
SPECIAL OPERATIONS	INSERT AND EXTRACT SPECIAL OPERATIONS FORCES
DRUG INTERDICTION	TRAIL COURIER AIRCRAFT AND BOATS, DEPLOY LAW ENFORCEMENT PERSONNEL
ANTISUBMARINE WARFARE	DETECT AND ATTACK SUBMARINES

The next mission examined was the Air Force special operations mission. In this mission, the aircraft are used to insert and extract special operations forces.

SLIDE 33



Our measure of effectiveness here is the relative number of special operations missions that can be completed each day. We looked at three alternatives: the V-22, the MH-53, and the H-47 helicopters. We also show our current capability.

The important variable in this portion of the analysis was the amount of darkness available to accomplish this mission. For obvious reasons, the Air Force would like to be able to carry out the special operations airlift under cover of darkness to maintain the covertness of those operations. So as the number of hours of darkness diminishes, the speed of the aircraft becomes more and more a factor in the ability to accomplish the mission. You can see that there is very little effect on the V-22 as the number of hours of darkness is reduced from 9 to 6 hours. The helicopter alternatives, given their substantially slower speed, would be able to complete far fewer operations in a limited period of darkness.

SLIDE 34:

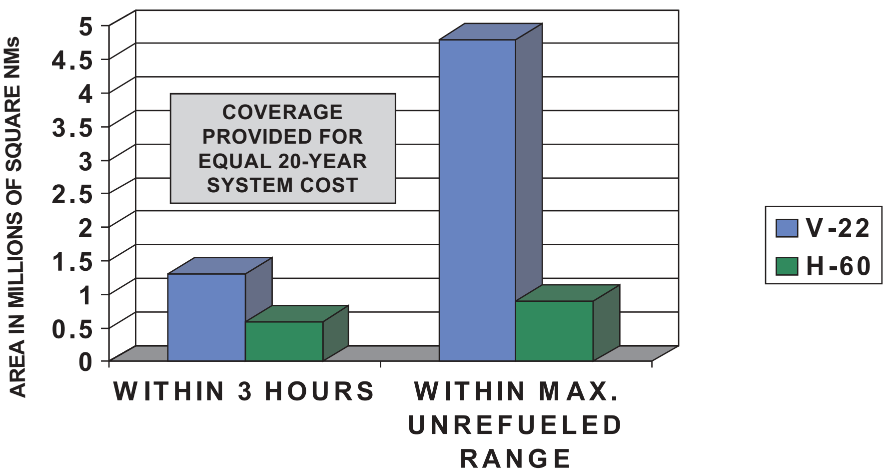
OTHER SERVICE MISSIONS	
MISSION	**CONTRIBUTION OF ASSAULT AIRCRAFT**
COMBAT SEARCH AND RESCUE	RECOVER DOWNED AIRCREWS
SPECIAL OPERATIONS	INSERT AND EXTRACT SPECIAL OPERATIONS FORCES
DRUG INTERDICTION	TRAIL COURIER AIRCRAFT AND BOATS, DEPLOY LAW ENFORCEMENT PERSONNEL
ANTISUBMARINE WARFARE	DETECT AND ATTACK SUBMARINES

The next mission examined was a drug interdiction mission that could be carried out by the Department of Defense or one of the other Government agencies. The aircraft would be used to trail the aircraft and boats that are used by drug couriers and to retrieve evidence and deploy law enforcement personnel.

SLIDE 35

DRUG INTERDICTION MISSION

- **AIRCRAFT CONFIGURATION**
  - **Air Surveillance Radar**
  - **Infrared Search and Track System**
  - **Turreted Gun or Two Machine Guns**
  - **Four-Man Crew**
  - **Five-Man Apprehension Team**
- **INTERCEPT CAPABILITY VERSUS LIGHT AIRCRAFT**
  - **Only Smallest Light Aircraft are Slower Than H-60 (160 knots)**
  - **Only Turboprops and Executive Jets are Faster Than V-22 (275 knots)**
- **COST-EFFECTIVENESS**



To carry out this mission we assumed the aircraft would be configured as shown here with a radar and an infrared search and track system. It would carry a four-man crew and a five-man law enforcement team.

We measured capability for this mission by looking at the ability of the aircraft to overtake the light aircraft that might be used by drug couriers. A helicopter with a dash speed of 160 knots would be able to overtake only the smallest light aircraft, whereas the V-22 with its dash speed of 275 knots would be able to overtake all but the larger turbo prop and executive jets.

In terms of cost-effectiveness, we measured the coverage that could be provided by the aircraft for equal 20-year system costs. The coverage that could be provided within a 3-hour time period is about 1.5 million square miles for a V-22 compared to about one half that number for a helicopter. The maximum distance in unrefueled range would be nearly 5 million square miles for the V-22 and only about 1 million-square miles for a helicopter.

NAVY'S ANTISUBMARINE WARFARE

SLIDE 36:

OTHER SERVICE MISSIONS	
MISSION	**CONTRIBUTION OF ASSAULT AIRCRAFT**
COMBAT SEARCH AND RESCUE	RECOVER DOWNED AIRCREWS
SPECIAL OPERATIONS	INSERT AND EXTRACT SPECIAL OPERATIONS FORCES
DRUG INTERDICTION	TRAIL COURIER AIRCRAFT AND BOATS, DEPLOY LAW ENFORCEMENT PERSONNEL
ANTISUBMARINE WARFARE	DETECT AND ATTACK SUBMARINES

The final mission that we examined is the Navy's antisubmarine warfare mission in which the aircraft are used to carry sensors that detect enemy submarines and then launch torpedoes to attack those submarines.

SLIDE 37:

RELATIVE CAPABILITY TO DETECT ENEMY SUBMARINES			
AIRCRAFT OCEAN	**NORTH ATLANTIC**	**NORTH PACIFIC**	**INDIAN**
SV-22	1.0-1.1	1.0-1.1	1.0-1.1
S-3B	1.3	1.4	1.4

An unclassified version of our results is illustrated here, where we show the relative capability of the aircraft to detect submarines. In this case the Navy currently does not plan to replace its S-3 aircraft until the late 2010-20 period. Thus, on an equal-cost basis, the current S-3 fleet would provide 20- to 30-percent more capability than could be provided by the SV-22 alternative.

[There is no SLIDE 38 in the Congressional Record.]

SLIDE 39:

NEAR-TERM COSTS FOR ALTERNATIVE FLEETS			
AIRCRAFT ALTERNATIVE	**COSTS INCURRED FY 1991-1997**	**NET PRESENT VALUE OR DISCOUNTED COSTS**	**YEAR MEB ASSUALT CAPABILITY ATTAINED**
V-22 - NORMAL PRODUCTION	**$13.1B**	**$16.3B**	**1996**
V-22 - SLOWED PRODUCTION	**7.7**	**13.0**	**1998**
NEW HELO	**6.6-8.7**	**11.8-13.0**	**1999**
CH-47M	**5.8-7.9**	**11.6-12.8**	**1997**
CH-60(S)/CH-53E+	**8.4-10.5**	**13.6-14.8**	**1996**
CH-46E+/CH-53E+	**8.3-10.4**	**13.3-14.5**	**1998**
PUMA/CH-53E+	**9.0-11.1**	**13.6-14.7**	**1998**
EH-101/CH-53E+	**9.6-11.7**	**14.0-15.2**	**1997**
ALTERNATIVE IN FY1990 PRESIDENT'S BUDGET	**5.2**	**—**	**—**

NEAR-TERM COSTS

As I mentioned in my opening statement, there is considerable concern about the near-term cost of the aircraft. We tried to take that into account in our assessment.

This is a duplication of one of the tables printed in our summary. We show the near-term costs for all of the alternatives that would be incurred in fiscal years 1991 through 1997, the net present value or discounted cost associated with the cost flow for each of the alternatives, and then the year that a typical operational capability could be attained—in this case a brigade-sized amphibious assault. At the nominal production rate that has proposed for the V-22 aircraft before the program was canceled, its cost would have been markedly higher than those for any of the alternatives, $13.1 billion planned over the 7-year period 1991 through 1997.

Given that, we constructed with the aid of the steering committee a slowed production alternative for the V-22 that would stretch the program out and procure aircraft at a slower rate than had been originally proposed. The original production rates were on the order of 40 to 60 aircraft per year. We slowed that rate to 36 aircraft per year. Doing that would reduce the near-term cost of the V-22 to just under $8 billion. Those costs would be in the range of those that we estimated for the helicopter alternatives. The alternative that was proposed in the President's fiscal year 1990 budget would amount to $5.2 billion. There are some reasons for those differences. I would be prepared to go into those during our question session if you would like. The principal effect of slowing the production rate for the V-22 would be to delay by 2 years the time that it could have an effective operational capability from 1996 until 1998. Some of the helicopter alternatives would provide an earlier capability: a CH-60 fleet, for example, in 1996, the CH-47 or the EH-101 by 1997.

SLIDE 40:

SUMMARY

MARINE CORPS MISSIONS

- GREATER SURVIVABILITY OF V-22 AND NEW HELICOPTER PROVIDES THEM A SLIGHT TO MODERATE ADVANTAGE IN AMPHIBIOUS ASSAULT MISSION
- V-22 IS MOST COST-EFFECTIVE ALTERNATIVE FOR SUSTAINED OPERATIONS, HOSTAGE RESCUE/RAID AND OVERSEAS DEPLOYMENT MISSIONS

ALTERNATIVES TO V-22

- NEW HELICOPTER DESIGNED TO MARINE CORPS REQUIREMENTS IF WILLING TO START NEW DEVELOPMENT
- MARINIZED CH-47 IF PROBLEMS WITH QUADRICYCLE GEAR CAN BE OVERCOME AT LOW COST
- COMBINATION OF SMALLER HELICOPTERS TO CARRY TROOPSAND CH-53ES TO CARRY HMMWVs

SUMMARY OF RESULTS

A concise two-slide summary of our results is shown next. For the Marine Corps mission, we found that the greater survivability of the V-22 and the new helicopter, which we assumed to be designed with survivability characteristics comparable to those of the V-22, would provide them a slight moderate advantage in the amphibious assault mission with a range of values typical of those that we showed you earlier.

For the other three Marine Corps missions, the sustained operations, the hostage rescue, and the overseas deployment, the V-22 was clearly the most cost effective of the alternatives.

In terms of the Marine Corps missions, the next best alternative would be a new helicopter specifically designed to Marine Corps requirements if the government were willing to start a new development program at this time.

After the new helicopter, the next best alternative would be a marinized version of the Army's CH-47 Chinook. There is a potential problem with its landing gear and the ability to use that aircraft aboard ship. It would be necessary to find a solution to that problem. We were not able to examine that in detail in our assessment. After those, the next best alternative would be a combination of any of the smaller helicopters such as the H-60 Blackhawk, to carry troops and the larger CH-53's to carry the marines' equipment.

SLIDE 41:

SUMMARY (Continued)

OTHER SERVICE MISSIONS

- V-22 IS MORE COST-EFFECTIVE THAN HELICOPTER ALTERNATIVES FOR SPECIAL OPERATIONS, SEARCH AND RESCUE, AND DRUG INTERDICTION MISSIONS
- S-3B IS MORE COST-EFFECTIVE THAN SV-22 FOR ANTISUBMARINE WARFARE MISSION

NEAR-TERM COSTS

- HIGHER PROCUREMENT COST OF V-22 LEADS TO LARGEST NEAR-TERM COSTS
- BY SLOWING PRODUCTION AND LENGTHENING TIME TO FIELD AIRCRAFT, NEAR-TERM COSTS FOR V-22 COULD BE REDUCED

For the other service missions, the V-22 was the most cost effective for the special operations, the search and rescue and the drug interdiction missions. Given the Navy's plan to retain the S-3B aircraft well into the post-2000 time period, that aircraft would be more cost effective than an antisubmarine warfare version of the V-22 aircraft.

As we showed you, the higher procurement costs and the near term funding requirements in terms of facilities and production line development would make the V-22 have the largest near-term cost of any of the alternative fleets. To some extent, we could mitigate those costs by slowing production and lengthening the time to field the aircraft.

At this time, Dr. Randall and I will be glad to answer any questions you have about our work.

Senator INOUYE. Dr. Simmons, I thank you very much. Your testimony has been extremely helpful.

I am certain all of us here are aware that this special hearing was convened because of the extraordinary concerns expressed by my colleague from Pennsylvania, Senator Specter. I believe he has spent more time than any one of us in studying the cost factors, the capability and effectiveness of the Osprey. Accordingly, I will yield to Senator Specter.

Senator SPECTER. Thank you very much, Mr. Chairman. Again, let me express my thanks to you, Mr. Chairman, for convening this hearing. I think that its importance was established when Secretary Cheney, in making a long list of proposed cancellations, put the V-22 Osprey at the top, No. 1, and I think that there have been very substantial concerns expressed in many quarters about the conflict between what Dr. Simmons has concluded and the summary rejection by the Secretary of Defense.

PANAMA

Dr. Simmons, I begin with the question as to General Pittman's statement that the V-22 would have saved lives in Panama. I would ask you whether you agree with General Pittman's conclusions that had we had the V-22, in fact, lives would have been saved in the Panama operation?

Dr. SIMMONS. Senator, I cannot answer that question directly because I have had no detailed knowledge of the Panamanian operation other than what has been discussed in the news or the print media. Certainly, the V-22 relative to the other aircraft that we examined specifically in our assessment would have considerable survivability advantages. A considerable amount of attention has been paid to the design of the V-22 aircraft that would give it a greater ability to survive hostile fire than would be the case in particular for some of the helicopter alternatives that we examined.

Senator SPECTER. With respect to a very fundamental point, there is no doubt that the V-22's landing ability would have prevented combat losses, deaths and injuries attributable to the parachuting operation.

Dr. SIMMONS. Again, Senator, I am not aware of the details, so I really cannot discuss or answer the question in that regard.

Senator SPECTER. Let me move to another high-profile item.

Senator INOUYE. Would the Senator yield? Maybe Dr. Randall can respond to that. Can you, sir?

Dr. RANDALL. I would be in pretty much the same situation, Senator. I think that is not an operation that we have analyzed at IDA and not something that we could comment on very directly.

Senator SPECTER. I will not press the Institute of Defense Analyses any further, Mr. Chairman. I think we might take congressional notice, or judicial notice, that if you can land in a helicopter, you can avoid death or injury from an airborne jump, and that those casual-ties could have been avoided. I think that attests to the conservatism and adds credibility to the testimony of Drs. Simmons and Randall when they conclude that the V-22 is a superior craft for the reasons that they have articulated. I start with General Pittman because it is on the record that the V-22 would have saved lives in Panama. I wanted to cover that point and I think we have.

HOSTAGE RESCUE

I was about to turn to another very high visibility operation, and that is the operation for the hostages in Iran in 1980. Fortunately, we have not needed another such rescue operation, but I think it is very important when we evaluate the V-22 and alternatives that very considerable attention paid to hostage rescue, as you have. While I realize that you have not made a specific study of the rescue operations in Iran in 1980, there has been considerable commentary on the failures of the helicopters and the difficulty of moving in the rescue operation. I would ask for your generalized opinion as to the significant, if not vast, superiority of the V-22 should the United States again be confronted with a similar rescue operation as that of 1980.

Dr. SIMMONS. A representative hostage rescue or small-scale combat raid was, in fact, one of the missions we examined in our assessment. As we showed there, the speed and range capabilities of the V-22 which, as you have noted, Senator Specter, are superior to those of the helicopters would, in fact, give an advantage in those types of operations.

CRITICAL CONCLUSIONS

Senator SPECTER. Dr. Simmons, I appreciate the testimony that you have given today and the very extensive report covering approximately 1.200 pages. I would summarize for purposes of this hearing your critical conclusion at volume I, page 13, and ask if you stand by this central and critical conclusion: "The V-22's speed, range, and survivability advantages could enable even the 356 aircraft fleet"—that is the number purchased with the smaller figure—"to be more effective, sometimes significantly more and other times only slightly more, than all of the proposed helicopter alternatives in each of the four Marine missions examined."

That, I believe, Dr. Simmons, is the central issue as far as the Marines are concerned. Is that accurate?

Dr. SIMMONS. Yes, Senator; that is the principal summary statement of our report, and we continue to stand behind that statement.

Senator SPECTER. The other critical summary statement appears in Volume I, page 15, and you have covered this in extenso but I think the crisp conclusion is important. I ask the same question about the accuracy of this conclusion. I will skip to Volume [?], page 22. "The V-22 is more cost effective than helicopter alternatives for the Navy combat, search and rescue, Air Force special operations, and DoD or other Government agency drug interdiction missions."

Is that precise and accurate?

Dr. SIMMONS. Yes, sir.

COST EFFECTIVENESS

Senator SPECTER. Coming to the issue of cost, Dr. Simmons, there is a statement in the letter of transmittal from Secretary Cheney that your report's V-22 program would cost about $3.7 billion more than the Secretary of the Navy has recommended for the Marine Corps medium-lift operation.

My question to you is whether that statement excludes the budgeted cost of the heavy lift in the special operations?

Dr. SIMMONS. Yes, sir; I think it does. As I understand it and perhaps you could seek clarification from Assistant Secretary Chu later on - no funds are currently provided in the DoD budget for either the CH-53 heavy-lift or the Air Force special operations aircraft.

Senator SPECTER. Could you give us an approximation as to what those costs would be?

Dr. SIMMONS. Yes, sir; the totals that we have estimated and, in fact, the totals we show in the final table of the executive summary of our report include the costs not only for the Marine

Corps medium-lift assault aircraft but the Marine Corps heavy-lift assault air-craft, the special operations aircraft and the combat search and rescue aircraft.

Senator SPECTER. What is the figure for heavy lift?

Dr. SIMMONS. The heavy-lift CH-53E helicopter, depending on the rate at which the aircraft would be bought over the 7-year period from fiscal year 1991 to 1997, would total from $1.1 to $1.7 billion.

Senator SPECTER. How about the special operations?

Dr. SIMMONS. Again depending on the rate at which aircraft are bought for that mission, the costs would be from $0.7 billion to $1.5 billion, all in the same constant fiscal year 1988.

Senator SPECTER. And what was the third category?

Dr. SIMMONS. Navy combat search and rescue.

Senator SPECTER. What would that cost be?

Dr. SIMMONS. $0.5 to $1.2 billion, again depending on the rate at which aircraft are bought for that mission.

COST DIFFERENCES

Senator SPECTER. You testified, Dr. Simmons, with respect to the $5.2 billion figure on the alternative in fiscal year 1990's Presidential budget, and you said you might have some additional amplification to give. Would you please do so at this time?

Dr. SIMMONS. Yes, Senator. There were three principal factors that account for the differences between our estimates of the cost for a CH-60/CH-53 fleet that would be comparable to the one that the Department of Defense substituted for the V-22. The first of these factors is that we elected to stretch the configuration - the fuselage of the CH-60 so that it could carry more troops at once. This increased the near-term costs for the CH-60 but reduced the overall cost for that portion of the alternative fleet. The second area in which costs differed were the avionics assumed in the aircraft. We found in our assessment that all of the alternatives benefited substantially by having avionics comparable to those planned for the V-22. The greatest benefit comes from increasing the ability of the aircraft to operate at night, particularly by the provision of the infrared night piloting system. This feature would make all of the aircraft substantially more survivable than they would be without that type of avionics. As a result, we elected to include the costs for those avionics in the summary table that we showed in the front of our report. The option of using less capable avionics in all of the aircraft, including the V-22, was also examined in the study and the same relative rankings obtained for the alternatives; namely, that the V-22 was slightly to moderately more cost effective than the helicopter alternatives.

The third factor that accounts for the difference in the costs are increases in the estimated procurement cost for the CH-53 heavy-lift helicopter. In the interval since, the Department's estimate for the costs of the CH-60/CH-53 fleet were prepared, the contractor that builds the CH-53 for the Government has indicated that the costs of that aircraft would be $5 to $7 million higher than was last reported to OSD.

So those are the principal factors that made our fleets more expensive than the alternative proposed by the Department.

COMPARABLE COSTS OF ALTERNATIVES

Senator SPECTER. Dr. Simmons, where you have projected the cost of the V-22 at the slow production, what would be the comparable cost of some alternatives? You have testified that in order to perform the functions of the V-22 you are going to need a combination of the CH-53's and the CH-60's; if you add those two together, what cost figure do you get in the comparable period?

Dr. SIMMONS. At the production rates we assumed, those costs would be as we showed in the summary chart. I believe the numbers were on the order of approximately $8 billion, give or take.

Senator SPECTER. For the alternatives, $8 billion, contrasted with $7.7 billion for the V-22 at the slow production rate.

Dr. Simmons, when the results of the Institute of Defense Analyses were disclosed, a statement was made on behalf of the Secretary of Defense, Mr. Cheney, by Mr. Pete Williams, saying that the Secretary disagreed with the assumptions of the Institute for Defense Analyses. Did not the Department of Defense have total opportunity to question your assumptions or to provide additional information at every stage of your proceeding?

Dr. SIMMONS. Yes, sir; I would say they did. As I mentioned in my opening statement, we interacted five times during the course of the study with our steering group, of which more than half the members of the committee were members of the Secretary's staff.

Senator SPECTER. Mr. Chairman, there are many, many more questions, but that is the distillation of the highlights. And my suggestion would be that if we hear from Dr. Chu, I would like to reserve the balance of the questions until Dr. Chu has a chance to present the essence of his conclusions, and then really get a more effective contrast of the expert judgments. So I would limit my questions at this time to the ones that I have asked, which are relatively minimal.

Senator INOUYE. Do you wish to have Dr. Simmons remain here?

Senator SPECTER. Yes, Mr. Chairman, I think that would be the better process, so that he could hear Dr. Chu. And where there are differences, we can crystallize them in the course of the hearing.

Senator INOUYE. Would you favor us with your presence?

Dr. SIMMONS. Yes, sir.

Senator INOUYE. Thank you very much, sir. I have a few questions here. Your study shows that the unit cost of the Osprey for the first 12 will be $123 million.

Dr. SIMMONS. I think one has to exercise some caution, Senator, in making the divisions from the cost numbers that we have shown. If one were to include that portion of the nonrecurring, one-time facilitization and production line setup expenses, then, yes, sir, that division would be accurate. We typically take the one-time costs out and set them aside to be added up later. And I believe what would be called the recurring costs for the aircraft would start out at around $80 million per aircraft, and then rapidly drop off to $30 million to $35 million. And, in fact, over a 356 to 400-aircraft buy, the average cost would be about $35 million.

CH-53

Senator INOUYE. In your study, it appears that the depot maintenance costs for the CH-53E are about 28 percent higher than that of the V-22. Considering that the Osprey is a bit more sophisticated than the helicopter, how do you explain this difference?

Dr. SIMMONS. Well, in estimating not only the depot level costs, but all of the operating and support costs for all of the aircraft, we used the same cost-estimating methodology in all cases. That methodology was one that had been originally developed by the Navy specifically to estimate the cost of rotary wing aircraft. The cost analysts who worked with me on the study obtained that model from the Navy, went through it in detail, collected all of the necessary inputs, and estimated the cost based on the characteristics reported for the aircraft.

I think one factor that needs to be taken into account is that in our analysis in the cost that you are describing, the V-22 and the CH-53 would be assumed to have the same avionics. So one of the principal differences between the V-22 and the current aircraft would have been removed in estimating those costs.

As to whether the V-22 is considerably more complex, I am not sure I would go that far. The two aircraft—the CH-53 and the V-22 —are approximately the same size within a few thousand pounds. Both of them are around 30,000 pounds empty weight aircraft. And I am not sure there would be that much difference in complexity. One of the factors that could be contributing to the difference in the depot level maintenance costs would be the fact that the CH-53 has three engines and the V-22 has only two. And engine costs could contribute a substantial amount to that.

Senator INOUYE. Why did you assume that all the aircraft considered would have the same avionics?

Dr. SIMMONS. We considered actually, Senator, a range of avionics in the aircraft. We wanted to eliminate from the argument the differences in capability that would result from differences in avionics. So at the extreme end we assumed that all of the aircraft would have avionics comparable to those of the V-22, which, as you know, has about $2.6 million per aircraft planned for the avionics suite.

As a percentage of the total aircraft costs that is a relatively small amount of average procurement devoted to avionics. It is something on the order of 8 to 10 percent.

As you know, some of our other more modern aircraft have substantially greater amounts of avionics. So we did not think that the V-22 type avionics was a dramatic departure in fitting that into all of the aircraft. But we went ahead and looked at other levels of avionics, leaving out some of the items that have been planned for the V-22 and seeing how the aircraft would operate with those.

As I mentioned earlier, one of the big differences was in the capability of the aircraft to operate at night. The infrared night piloting system planned for the V-22 and included in that set of avionics provides substantial night operations capability that results in greater survivability for the aircraft. And those results were apparent in the effectiveness results that we calculated for the amphibious assault mission.

Senator INOUYE. Dr. Simmons, as I have indicated, if you could step aside for a few minutes, we would appreciate it very much.

DEPARTMENT OF DEFENSE

STATEMENT OF DAVID S.C. CHU, ASSISTANT SECRETARY OF DEFENSE FOR PROGRAM ANALYSIS AND EVALUATION

Senator INOUYE. And now, may I call upon Dr. Chu, the Assistant Secretary.

David Chu, Assistant Secretary of Defense for Program Analysis and Evaluation.

Secretary Chu, we are very happy to have you with us, sir.

Dr. CHU. Thank you, sir. Good morning. I do have a longer statement, which I would like to submit for the record if I may.

Senator INOUYE. It will be made part of the record.

Dr. CHU. Thank you.

And I would attempt very briefly to summarize the principal points behind the Department's position on this aircraft, particularly in light of the IDA study, before responding to your questions.

COST CONCERNS

As Mr. Cheney has stressed, the fundamental difficulty with the V-22, as far as the Department of Defense is concerned, is the high up-front investment costs that are involved. If you translate the costs in the IDA report into today's dollars, fiscal year 1990 dollars, and look at the 356-aircraft buy averaged over the life of that program, this aircraft will cost us on average about $42 million apiece. The hope from the beginning with this program has been that one could amortize the high up-front investment costs with smaller operating costs at some future point. Speaking frankly, the Department's track record in estimating future operating costs of new systems is not splendid. And in general, people tend to underestimate, even with the best intentions and the best models, what future operating costs are going to be.

The difficulty here is you are comparing a new system and its unknowns and uncertain-ties, the V-22, or a new helicopter, which was also looked at in the IDA study, with a series of systems with which the Department has a great deal of experience. And there is great doubt that one would ever realize the kinds of operating cost savings necessary to offset the up-front investment costs. Even if one did, it would be well beyond the year 2000, perhaps 2010 or 2015, before you saw a break-even point, and ignoring the fact that one is always more worried about money today than money tomorrow.

Perhaps more to the point, as Mr. Cheney stressed in his testimony to you just last month, the department's difficulty, as you all appreciate at least as well as we do, is how to manage in the decade of the 1990's with substantially less funds than it has previously enjoyed; how to do reasonably well in that fiscal climate with its most important missions; how to avoid letting "better" be an enemy of "good enough."

We have to do reasonably well, not perfectly, with these major missions with a lot less money than we have had before, and a lot less money than people had previously planned on when this aircraft development program was launched in the early 1980s. And, in fact, as Mr. Cheney stressed in his testimony, relative to the fiscal profile that pertained last year when he made his decision to cancel this program, the Department's top line in terms of the President's recommended numbers for fiscal years 1990 through 1994 has fallen by $167 billion, the equivalent of roughly one-half to two-thirds of a year's budget for the Department.

And in fiscal year 1991 alone, $22 billion has been removed. That top line, as you all appreciate, is substantially higher than the Congress is likely to enact for the Department. The Senate Armed Services Committee has, as we are all aware, removed approximately $18 billion from the fiscal year 1991 number already in its mark.

Therein lies the great difficulty with proceeding with this program. As the Secretary stressed in his letter, relative to what the Secretary of the Navy would recommend we spend on the medium-lift requirement, setting aside heavy-lift and other missions as separate issues, you would need almost $4 billion more in the period 1991 through 1997 than the Navy is prepared to spend on this mission, even at the President's fiscal numbers, which are higher than the Congress is likely to sustain in the end. Moreover, it would take you to about 2009 to finish buying the aircraft, even for the smaller IDA buy, and at the slower procurement profile that IDA has identified in its summary report.

What this would compel the Department to confront are a series of very painful tradeoffs to find the several billion dollars necessary to sustain that buy, not only in the period 1991 through 1997, but in the years beyond. The Secretary specifically pointed to the clear tradeoff within the amphibious mission area, which is whether we have enough money to buy both the ships that the Marines need and the aircraft, if we go for an elegant aircraft solution.

I think the essence of the Secretary's decision a year ago and his reconfirmation of that decision today has been to try to preserve sufficient funds so we could have a reasonable amphibious lift shipbuilding program, and a reasonable helicopter lift program for the Marines. If we instead select the V-22, what we are doing is driving some other important element out of the Department's budget or limiting our ability to afford it. And one of the things the Secretary has pointed to that we would have to consider is whether we can afford the kind of amphibious lift that we would need to get the troops to the theater of operation in the first place, however elegant the method then used to get them ashore.

STUDY ASSUMPTIONS

Now let me turn to the study itself. As in any study, as I think the questions here have elicited already, the conclusions are a function of the assumptions that one makes. And I think there are three broad assumptions or sets of assumptions that are useful to focus on in thinking about the conclusions this study has reached.

V-22 PRODUCTIVITY

One is that for the V-22 itself, what IDA has done in its smaller buy, the 356-aircraft buy, is, in essence, assume much higher productivity from the V-22 than the Marine Corps itself has heretofore been willing to assume in positing landing requirements in its studies.

A key element in that productivity is flying the aircraft at fairly high speeds with external cargo loads. There had been some doubts earlier about whether that was wise or feasible. There still, in fact, remain some things that have to be tested and proven out. And obviously, there is the doctrinal question of whether the Marines want to move to the kind of sortie rate implied by the 356-aircraft fleet. It is a higher sortie rate than even their most recent amphibious lift study has identified as being their standard.

HELICOPTER PRODUCTIVITY

On the other hand, as far as the helicopters are concerned, I think the study limited helicopter productivity relative to their potential by a number of different assumptions. One of them is this question of whether you can carry, and to what extent can you carry, two vehicles externally as cargo under the larger helicopters, the CH-53 and the CH-47, or what is called dual-sling loading in the study.

I think some perspective on this issue is helpful. A year ago at this time, although the concept was tested in 1985 by the Marine Corps and viewed as promising, based on the limited tests that were done at that time, dual slinging was viewed as a heretical proposition. In the year that has elapsed, the Marines themselves in their latest lift study have come to the conclusion that one-half the time one could dual-sling load such cargo and the other one-half of the time carry only one vehicle per helicopter.

This is an important assumption in differentiating the productivity of the helicopters from that of a V-22 force. The larger helicopters could carry two vehicles apiece. To the extent they can carry two more frequently, they would be more productive, and you cut the size of the helicopter fleet that you need.

A second assumption that limits helicopter productivity is the issue of how many troops are carried internally in a helicopter. To take an example, as I understand the study at least, typically, when the CH-47 is used in the study, only 24 troops are carried internally. The helicopter is capable of carrying 42 troops.

Why were 24 assumed?

That is the Marine Corps standard, and it was not challenged in the study.

A third assumption that is important in the results is the doctrinal view that troops move first, then vehicles. This is an assumption that I think, on a net basis, disadvantages helicopters vis-à-vis the V-22. You basically have a certain amount of helicopter capacity waiting in the all-helicopter cases until all the vehicles are moved.

A fourth assumption that I think has an important effect on the results is the use of what is called the HMMWV, the high-mobility multipurpose wheeled vehicle, which is the small truck to which the Marine Corps has moved for its payloads. That, in essence, is the big element in cargo movements from ship to shore.

If one were willing to consider going to a smaller vehicle and downsizing some of the cargo payloads that are now put on that truck, which the Marine Corps is examining, to be fair about it—it is currently debating an operational requirement for just such a vehicle—such a vehicle could fit inside many of the aircraft, now being discussed, as opposed to being carried externally. This would have a significant effect on the cargo-carrying capability of the different alternatives. I think it would reduce substantially the size of the helicopter fleet one would have to have in order to move the same force from ship to shore.

Finally, we would have some quarrel and this is really an empirical matter about which we are still engaged in dialog with IDA regarding the reliability assumptions that are made in the study, vis-à-vis the helicopters. I think we would argue they are somewhat more reliable than they have been assumed in the study to be. That also affects the kind of results that one gets from the study.

DISCOUNTING

The third broad class of assumptions to which I would call attention is the focus in the study on undiscounted costs. The alternatives were developed before any attempt was made to even out the differential time streams by applying discounting. So that one does not get the same appreciation of the cost problem as one does by focusing on the large upfront, near-term investment costs associated with the V-22, as opposed to the hoped-for long-term savings that it might produce.

And as a further element of that, one does not, I think, as much consider the uncertainty that is attached to the operating support costs of the V-22 as ought to pertain.

SUMMARY

To sum up, the Secretary has stressed repeatedly, I believe, that he views the V-22 as a potentially interesting aircraft. He is not criticizing the airplane. It is simply a matter that the Department cannot afford the near-term, up-front costs of buying this aircraft. And I think it is important to put its attractiveness and its many fine features into that perspective.

I think one of the very interesting conclusions in the IDA report is that all the alternatives, including the helicopter alternatives, generally perform better than the current force, often substantially better. And so this is not an issue of whether the Department is going to upgrade medium lift for the Marine Corps. The debate is over the best way in which to do that what is the best way, in particular, in light of the very severe fiscal constraints the Department faces in the coming years.

Thank you, Mr. Chairman.

[Dr. Chu's prepared statement follows:]

STATEMENT OF DR. DAVID S.C. CHU

Mr. Chairman, Members of the committee:

Good morning. I am David Chu, the Assistant Secretary of Defense for Program Analysis and Evaluation. I am here at the request of the Subcommittee, and I will do my best to answer your questions concerning the Department of Defense's position on the V-22 program in light of the draft IDA study report that has been provided to you.

To summarize the Department's position: The V-22 program even the scaled-down version in the IDA study remains unaffordable in today's budgetary climate, which is likely to become even more stringent. While the V-22 has many positive attributes that no existing helicopter can match, it is still a tactical transport that would cost about $42 million per copy in today's dollars.

The major problem with the IDA study is that it attempts to optimize the use of resources at levels much higher than those now planned for the missions it examines, whereas the Department's problem is to find ways of performing the most critical missions at acceptable levels with considerably smaller resources. More specifically:

- The up-front costs of the V-22 significantly exceed those of the helicopter alternatives. (IDA does not seem to disagree on this basic point.)
- Offsetting long-term savings associated with the V-22 option are dubious and would accrue gradually over many years to the extent that they occur at all.
- The alternative helicopter forces perhaps in smaller numbers at lower investment and operating costs can perform the higher-priority DoD missions at acceptable levels.
- In fact, all of the helicopter forces examined by IDA do better than today's force: thus the issue is how much more performance we should seek.
- Viewed in that broader context, the V-22's acknowledged advantages diminish in attractiveness.
- And the V-22 versus helicopter comparison is inapplicable for two of the missions treated in the study because DoD does not currently plan to buy mare aircraft of any type to support these missions.

The attached paper elaborates on these points and provides additional material on the DoD position. In this paper we have generalized for brevity and to provide an unclassified discussion. We have attempted to do so fairly.

I would be delighted to answer your questions.

DOD POSITION ON V-22 IN LIGHT OF THE IDA STUDY

Introduction

- The Institute for Defense Analyses (IDA) study of the V-22 and helicopter alternatives was directed by the Congress; the Congressional tasking also mandated coverage of several potential V-22 missions that had not previously been considered.
- The study provides a lot of new information on the cost-effectiveness of the V-22 versus helicopter alternatives, but it does not support reversal of Secretary Cheney's decision to cancel the V-22 program on the basis of affordability.
- The decision to terminate the V-22 was made in the context of an overall Defense program that was considerably larger than the FY92-97 program now being developed by DoD, which many In Congress argue should be reduced further.
- The principal shortcoming of the IDA study is that it focuses on optimizing the use of a resource level that is far higher than DoD will receive. DoD's challenge is to figure out how to accomplish its most important missions—perhaps less well than would be possible with more resources—while achieving substantial savings.
- Moreover, the IDA study assumes we can change operational practices to play to the V-22s major strengths but downplays the advantages of the larger helicopters in the alternative forces. For example:
- In its primary sizing case—opposed Marine Corps helicopter assault—the V-22 flies at a sortie level almost 50% higher than the rate posited in the Navy's most recent amphibious lift requirements study. To achieve this sortie level, V-22 must fly at speeds

approaching 200 nautical miles per hour, with externally carried loads, without damage to that cargo.

- The substantially greater payload of the CH-53E and CH-47M helicopters at ranges out to 200 or 300 NM is given relatively little credit (e.g., the CH-53E requires one-fourth fewer sorties to move HMMWVs, although this does not change the number of CH-53s bought for IDA's equal-cost forces).
- While these large helicopters are more vulnerable than the V-22 to enemy small arms fire, it should be noted that we already have CH-53s for demanding combat missions Marine Corps helicopter assault and special operations behind enemy lines and we are also buying MH-47Es for special operations.

- The remainder of this paper deals first with near-term and long-term cost aspects of the IDA study: it then discusses separately seven of the mission areas covered in the study: and it concludes with some brief observations concerning commercial applications of the V-22. In summary form, the following key points are made:
    - The up-front costs of the V-22 significantly exceed those of the helicopter alternatives.
    - Offsetting long-term savings associated with the V-22 option are dubious and would accrue gradually over many years to the extent that they occur at all.
    - The alternative helicopter forces perhaps in smaller numbers at lower investment and operating costs can perform the higher-priority DoD missions at acceptable levels.

Near-Term Costs

- The driving factor underlying the V-22 cancellation decision was the comparative "up-front" investment cost of the V-22 versus an alternative force of helicopters that was judged capable of performing two of the three military missions posited for the V-22 - Marine Corps helicopter assault and Navy search and rescue—reasonably well. (The third mission, extraction of special operations teams at ranges beyond helicopter capabilities, was simply eliminated as unaffordable.)
- The previously noted higher V-22 sortie rate (coupled with lower assumed peacetime attrition and modestly offsetting changes in lift requirements) results in a much smaller Marine Corps V-22 force than was under consideration when the program was terminated last year (356 versus 552 aircraft, a decrease of 36%).
- Even this downsized V-22 program would cost considerably more to buy in the up-coming program period than the same fraction of the equal-cost helicopter fleets developed in the IDA study.
- Over FY91-97, the IDA study posits buying 86 V-22s for Marine Corps missions (24% of the Marine Corps total of 356) at a cost of about $5.6 billion (FY88 $).
- The cost to buy the same 24% of the study's larger CH-47M, CH-53/60 and CH-53/46 fleets would range between $2.7 billion and $3.3 billion.

Long-Term Costs

- In the IDA study, the up-front costs of the V-22 are offset by the larger operating costs of the alternative helicopter forces over roughly the next 20 years.

- V-22 operating costs per aircraft assumed in the IDA study are slightly higher than those of the helicopters, as shown below (in FY88$)
  - $2.28 million per year for the V-22
  - $2.13 million per year for the CH-47M helicopter
  - An average of $2.14 million per year for a mix of CH-53 and CH-60 helicopters
- Helicopter force operating costs are increased in the IDA study through buying helicopters in considerably large numbers.
  - To some extent this is an artificiality of the IDA study's "equal cost" methodo1ogy—it keeps buying helicopters until the sum of acquisition and operating cost is even.
  - Whether helicopter forces of the resulting size would be needed is unclear, as discussed later.
- Two other points should be noted on this approach to determining force costs:
  - Future operating costs of the V-22 are more uncertain than those of the better-known helicopters. Whether the V-22 force will really cost less than IDA's much larger helicopter forces is simply unknown at this time.
  - Even if V-22 force operating costs are considerably lower, the savings thus generated would accrue in small annual increments that would not "break even" until 2015-2020. The annual operating cost for IDA's Marine Corps V-22 force of 356 aircraft would be about $530 million in FY88 dollars; IDA's 527 CH-47Hs and its mix of 525 CH-53 and CH-60 helicopters would both cost about $730 million per year. (Operating costs are only computed for 65% of each force that is actually operational; the remaining 35% are for "pipe-line" and peacetime attrition.)

Marine Corps Aerial Assault

- The MEF-size amphibious assault constitutes the most important prospective V-22 mission.
  - It determines the size of the total Marine aircraft buy (other elements are essentially developed proportionally).
  - And the Marine buy was 84% of DoD's previously planned V-22 acquisition (552 out of 657 aircraft).
- If one accepts as inviolate statements of lift requirements and aircraft load restrictions, and if one credits the V-22 with roughly 50% greater productivity than helicopters due to its speed/sortie rate, then the sizes of equal-cost helicopter fleets developed in the IDA study are correct.
- But if one changes productivity and requirements/loading assumptions the V-22 and helicopter forces converge in size.
- In fact the IDA study identifies the CH-47M as the best existing helicopter alternative to the V-22.
- This aircraft, although slower than the V-22 by a factor of 30 to 50%, is also capable of carrying payloads roughly 50 to 70% greater than the V-22 out to ranges of around 300 NM (allowing a round trip at the longest amphibious force standoff distance treated in the study, 100 NM).

  - The CH-47H can carry up to 42 troops, whereas the V-22 can only carry 24.
  - The CH-47H can also carry the High Mobility Multi-purpose Wheeled Vehicle, the dominant equipment item moved by the V-22 and alternative helicopters and 24 troops besides; the V-22 cannot do so.
  - If the HMMWV were replaced by a smaller vehicle, it could be carried internally by the CH-47H, producing greater mobility and firepower early in an assault. The V-22 can also carry a small HMHWV internally, but with only 11 troops vice the desired load of 24. (Aircraft carrying externally slung loads are more vulnerable as they must hover longer prior to landing; since the current HHMWV must be carried externally by all candidate aircraft, a potentially significant combat advantage could accrue to aircraft that can carry a smaller vehicle internally.)
- If one were to fully exploit the CH-47M' s larger troop-carrying capacity by moving 42 versus 24 troops in the initial assault phase of the operation, IDA's CH-47M force could be reduced by about 8%.
- If in addition one assumes that all HMMWVs can be carried in double loads (e.g., a smaller version internally and a regular one externally), then the CH-47M force could be cut by about 40%.
- If one instead assumes that only 30% of the aerial assault force's HMMWVs are converted to smaller versions that can be carried internally by the CH-47M along with 24 troops—to provide greater mobility and firepower early in the assault phase—then the CH-47H force could be cut by about 20%
- Another way of reducing helicopter loads would be to "double-sling" HMMWVs from the larger helicopters. (This concept has been tested on the CH-53 but not on the CH-47M, although the latter aircraft has the load-carrying capacity. The IDA study's base case assumes double-slinging of half the CH-53E's HMMWV loads.)
- While the IDA study suggests that the CH-47M is the existing helicopter that would cost least for approximately equivalent aerial assault capability, it is also likely that more flexible usage of the CH-53E and/or adoption of a small HMMWV would produce CH-53/60 and CH-53/46 helicopter combinations that could perform the aerial assault mission acceptably with smaller numbers than IDA computed on an equal cost basis.
- Regardless of whether we buy V-22s or helicopters for Marine missions, the "small HMMWV with troops" option should be examined carefully. (In the IDA study this concept only appears as an excursion.)

Sustained Marine Corps Operations Ashore

- "Requirements" for tactical airlift in support of combat operations are highly scenario dependent and therefore inherently soft.
- The IDA study's examination of this mission area is based on an Army study that assumed operations over a much larger area than would be typical for the Marine Corps following a MEF-sized amphibious assault.
- If larger forces and areas are involved, it is reasonable to assume tactical airlift support by Army helicopters and Air Force C-130s—forces we plan to have in any event.
- As shown in the IDA Study, the Marine helicopter forces perform fairly well in more confined areas.

    - Out to radii of operation of around 150NM, for example, the CH-47M has 50 to 70% better troop and cargo-carrying capacity than the V-22, which could be important in some situations. The slower speeds of the helicopters are less important over short distances.
- A critical factor driving the greater productivity of the V-22 is its higher utilization rate and ability to carry relatively small payloads over longer distances due to its speed. Again, this advantage shrinks as areas of operation become smaller.
- The study seems unduly pessimistic about the assumed utilization rate of the helicopters. For example, it employs the CH-47M at a rate of 1.7 hours per day, whereas Army data for the CH-47D suggest that a rate of 4 hours per day could be achieved.

MEU/SOC (Marine Counter Terrorist) Operations

- For this mission the IDA study looks at a specific scenario involving rescue of 50 hostages from a location within range of a Marine force embarked in amphibious ships.
- In this case, the V-22's speed, longer range, and assumed greater reliability confer advantages on the order of 5 to 7 hours in response time, which the study points out could be critical to mission success if the terrorists were to discover that the hostages' location had become known.
- Situations where this extra degree of responsiveness would be critical could of course occur, but they might not occur very often.
    - It is much easier to foresee cases (assuming any intelligence is available in the first place, the real problem here) in which neither a V-22 or helicopter equipped Marine force could respond in time, or in which both forces could.
    - If one makes less pessimistic assumptions about helicopter reliability, the response time advantage of the V-22 over the CH-47M drops to about 2 hours. If one further assumes random location of the amphibious group within 2 days steaming time of the hostages' suddenly-revealed position, and that use-ful intelligence expires linearly over the two day period, then the overall probability of mission success is on the order of 3% greater for a V-22-equipped force than for a CH-47M-equipped one.

Deployment Capability

- The IDA study points out the V-22's capability to "self-deploy" to considerably greater ranges than the alternative helicopters can achieve. While this is an acknowledged V-22 advantage, it should be put into context.
- The V-22 force does not entirely "self deploy;" at extremes of range it cannot carry any of the additional ground support equipment needed to operate, which must be airlifted by other aircraft.
- The overall difference in strategic airlift required to move a MEB-size force equipped with V-22s versus helicopters is small. For example, the percentage of projected airlift capacity generated over 20 days in moving to materiel pre-positioned on MPS shipping would be:
- 1.66% for a V-22 equipped MEB
- 1.77% for a CH-47H equipped MEB

- 2.04% for a MEB equipped with a mix of CH-53s and CH-46s

Special Operations

- DoD had planned to buy 55 V-22s for special operations, largely for long-range extraction of SOF personnel. (The V-22 can make such pick-ups at distances 40 to 70% beyond the capabilities of the helicopter alternatives; the HC-130 can insert and re-supply Special Forces at considerably longer ranges than either helicopters or the V-22 could retrieve them.)
- In this case, the decision was made simply to forego the capabilities brought by V-22 to this mission, relying on shorter-range helicopters already bought, or being bought, for SOF airlift, together with the longer-range HC-130s. Major savings on both acquisition and operations resulted from this decision.
- With the sizable investment already made, or planned, to build up SOF airlift capabilities, DoD can meet a substantial majority of currently postulated requirements.
  - Due to changes in world conditions and US force planning concepts, these requirements—particularly for long-range operations—are being re-evaluated.
- In the absence of firm, critical needs for extracting special operations forces in the range interval made feasible by the V-22, DoD's decision not to spend the IDA study's nominal 16 billion (or any other large amount) for either the V-22 or a lesser-capability helicopter force remains sound.
- In fact, IDA's results could call into question our decision to procure HH-47s; the study suggests that we might need a greater proportion of aerial tankers relative to helicopters.

Search and Rescue

- When the V-22 program was terminated, 50 H-60 helicopters were substituted for the 50 V-22s programmed for the Navy's search and rescue (SAR) mission (a variant of the H-60 is used by the Air Force for this mission).
- Considerable savings also resulted from this decision, although the loss in SAR capability was clearly recognized.
- While the V-22's SAR capabilities are superior in the cases examined in the IDA study, in other situations the difference could be less pronounced, e.g., when search speed is reduced to account for weather or light conditions.
- Overall Navy priorities for keeping carrier force levels up, and combat aircraft inventories modernized, effectively preclude expending additional resources to re-institute a V-22 buy for the SAR mission.

Drug Interdiction

- This is another case in which IDA did the best it could in examining V-22 or helicopter use in a new mission area, but appears to have focused somewhat too narrowly.
- DoD does not buy any aircraft explicitly for drug interdiction; instead it assists other government agencies in counter-drug operations, employing ships and aircraft procured for military purposes.

Commercial Applications

- Some supporting the V-22 cite its potential to give US industry the dominant role in what is asserted to be a lucrative commercial market, thus generating thousands of jobs and improving the U.S. balance of trade.
- This may or may not be an accurate forecast, but it should be noted that DoD has already, funded V-22 R&D amounting to more than $12.3 billion; it is not clear that further subsidy is needed.
- Moreover, if there were a profitable commercial market, presumably the V-22 consortium would stake its own money on behalf of such an enterprise. One has to believe that Boeing, a partner in V-22 and a world leader in aviation, with a proven track record of judging which are the profitable investments, would fund a commercial development if it were indeed an attractive proposition.

[End of prepared statement]

Senator INOUYE. Thank you, Mr. Secretary. Senator Specter.

Senator SPECTER. Thank you, Mr. Chairman.

SURVIVABILITY

Secretary Chu, do you disagree with General Pittman's conclusion that the V-22 would have saved lives in Panama?

Dr. CHU. I am not aware of his statement on that matter, sir. And I think I would be careful not to get into what could be construed as an inter-service rivalry on who was best equipped to carry out that particular mission.

Senator SPECTER. Well, the absence of General Pittman would certainly facilitate no inter-service rivalry.

Dr. CHU. Excuse me, sir?

Senator SPECTER. Keeping General Pittman away from these hearings helps out.

Dr. CHU. I am not quite sure what you are referring to.

Senator SPECTER. Let me begin again.

I represent to you that General Pittman said that the V-22 would have saved lives in Panama. Do you disagree with that conclusion?

Dr. CHU. I am not sure I could assess that conclusion without looking at the details of the actual operation and what kind of aircraft were or were not used. The V-22, of course, was not available in that particular timeframe.

Senator SPECTER. Are you aware of the loss of lives and injury in Panama, which resulted from parachuting operations?

Dr. CHU. Yes, I am.

Senator SPECTER. Would the V-22, which eliminates the necessity of parachuting, have saved lives and injuries?

Dr. CHU. Not necessarily. I think that is an important point. It raises a very interest-ing question. I think it is important to stress that, as the IDA study shows, all aircraft take losses in these operations. The issue is what is the relative loss. I would argue the differences are modest. But you are not going to conduct these operations without losses.

And so the issue you are raising is, would we have fewer losses if we used an aircraft approach, specifically V-22, as opposed to paratroopers. Without some careful calculations, and I am sure some quite energetic inputs from the Army, I would be unwise to speculate on which one is the better approach.

Senator SPECTER. So you say there would be modest differences, but differences nonetheless?

Dr. CHU. No, sir; that is not what I said.

Senator SPECTER. Wait until I finish the question, Dr. Chu.

Dr. CHU. Excuse me.

Senator SPECTER. And I will not interrupt your answer.

So you say there would be modest differences, but necessarily some differences, in loss of life and injury if you had the V-22, contrasted with not having it?

Dr. CHU. No, sir; what I said was that the IDA study indicated that you would have losses with all aircraft. Essentially, in terms of what the study focused on, the only difference in losses were to small arms fire. I would view the size of those differences as modest.

They are a function of whether the operations are at night or in daytime. The differences are much reduced at night. They are a function of the kind of terrain. They are a function of the specific assumptions IDA made about the altitude at which the aircraft flew. IDA did not, in fact, have the helicopters flying nape-of-the-earth operations, which would presumably reduce losses when carrying troops. The differences are a function of the reaction time one assumes for small arms fire.

My statement goes to what IDA found in its report. To the larger question you are raising, would we be better off eliminating paratroopers from the nation's military force structure and buying the V-22, I think I would defer in trying to speculate on what the answer to that question would be.

Senator SPECTER. All right. Apart from the obvious conclusion that there are no parachuting losses if a helicopter lands the troops, let us go to the report of the Institute of Defense Analyses specifically and pick up the comment on page 13:

The large CH-53E and CH-47M helicopters are 1.7 to 3.5 limes more likely to be downed by enemy air defenses than is the V-22.

Do you disagree with that very important finding?

Dr. CHU. I think I disagree with the implication you would like to draw from it, sir. No one disagrees with the specific calculations IDA has made. The absolute numbers are classified. I think when one looks at the absolute numbers, one puts these proportional differences into perspective. I think it is also important to stress that those numbers are a function of the assumptions made in the study.

Specifically, in the study, if I understand it correctly, the reaction time assumed for enemy small arms fire is right at the edge of the speed envelope of the V-22, such that, in fact, the enemy gunners do not have much chance against the V-22, and that accounts for its small losses. Assume a faster reaction time and the helicopters will look better relative to the V-22. Assume a slower reaction time and everybody is going to look great.

The other thing I think important to stress in this debate is, as the study indicated, that all aircraft in these operations are going to take losses. The CH-53, as I think you appreciate, sir, is going to be used in these operations in any event, whatever decision is made on the V-22. In the Marine Corps' earlier calculation of requirements, just under 30 percent of all sorties have to be flown by the CH-53 because it is the only aircraft either it or the CH-47 that can carry the heavier loads, the artillery, the larger trucks, the light armored vehicles that in earlier landing scenarios the Marine Corps planned to use.

So the CH-53 is going to be present, and we are going to endure losses for that aircraft whatever choice is made on the V-22. And the other point that I think is important to stress is we will have losses for the V-22 as well. It is not invulnerable.

Senator INOUYE. Will the Senator yield?

Senator SPECTER. Yes; I need time to reread the answer in any event, Mr. Chair-man.

Senator INOUYE. I gather that you may have classified information, which could materially affect the outcome of our decision. If you wish to submit such classified information we would be very happy to receive it in our secure facilities.

Dr. CHU. I would be delighted to, sir.

Senator INOUYE. Senator.

Senator SPECTER. Well, I am not going to pursue the question, except to restate the IDA conclusion about the high level of downing of the 53's and the 47's. And no one denies that there is a V-22 problem, but the comparison is very plain. There are so many questions to be covered that I think it not advisable to reanalyze your answer in terms of the question.

PRODUCTIVITY

You have testified, Secretary Chu that the higher productivity assumed by the Institute of Defense Analyses is higher than the Marine Corps has been willing to assume. Let us cut right to the core of the issue to General Gray, the Commandant's position, summarized in the Navy Times on March 5, 1990, "Gray: The V-22 substitute scheme is ridiculous."

Now the Marines have testified on many different occasions about their very, very, very, very, very strong preference for the V-22. Do you disagree with General Gray that all of the proposed alternatives, all the proposed substitute schemes are ridiculous?

Dr. CHU. Yes, sir; I respectfully do. After all, we are living with one of those substitutes—the least attractive alternative in the IDA study—today, and that is the current fleet. I think one of the very important points the IDA study makes is that the current fleet really ought to be replaced. Then the debate begins on what the replacement should be. What Mr. Cheney endorsed a year ago at this time was an alternative that had long been viewed as the Marine Corps', if you will, fallback position if it did not get the V- 22. That is, a combination of CH-60's—a fairly modern helicopter, with the survivability advantages of more modern helicopters, and that the Army is buying today and using today, and plans to use for a long time to come for combat assault—and additional CH-53's to carry the heavier materiel, specifically the HMMWVs.

There may be other alternatives that are just as good in the helicopter class, just as good or better than that one. In fact, I think the one, at least to my reading of the IDA report that comes out as the alternative that clearly merits further consideration is some variant of the CH-47. This aircraft, as I know you are aware, is being pursued today for use by the special operations forces in the form of the MH-47E variant, with highly sophisticated avionics aboard.

This aircraft, in fact, has the advantage of a very large payload, not as fast as the V-22, not as modern in terms of survivability enhancements as a CH-60, but it is a very attractive alternative, at least as presented in that study. So I would have to disagree with General Gray's view. I, more keenly than anyone, I think, appreciate how much the Marine Corps would like to have this airplane. Mr. Cheney's decision is simply that we cannot afford it in this decade.

V-22 COSTS

Senator SPECTER. Secretary Chu, the transmittal letter from the Department of Defense of the IDA report states that the smallest V-22 program contained in the report would cost about $3.7 billion more than the Secretary of Navy has recommended for the Marine Corps medium-lift aviation. It is difficult to come to a precise minimal characterization, but I would suggest that this is misleading, if not disingenuous. Because there are other functions, such as the heavy lift and special operations, which are comprehended within the V-22 mission. So when you put a figure on a transmittal letter which says that it is $3.7 billion more than the Navy has budgeted, you left out very critical aspects of what the Navy budgets in other lines.

Now, you heard the testimony of Dr. Simmons that when you add in the other functions you come to the range of $2.3 billion to $4.4 billion, so that, in fact, the budgeting of the Navy is just about the same for the V-22 as the other missions' budgeting, and is in fact nowhere near a $3.7 billion differential. How do you respond to the modest characterization of that statement as misleading or disingenuous?

Dr. CHU. I do not think I said it was misleading or disingenuous.

Senator SPECTER. No; I said that.

Dr. CHU. I apologize then, sir.

Let me stress, I think you misunderstood the Secretary's statement. The statement in the letter is an apples-to-apples comparison that compares only the V-22's that are bought for the medium-lift requirement with the funding that the Navy has proposed for that requirement.

If we want to add back in amounts for special operations forces (SOF) and for other missions, in terms of what the Navy's recommended amounts are, frankly, that is only going to widen the differential that was stated in the Secretary's letter. The bottom line here, sir, with great reluctance by the Department, is, we cannot afford to spend the kind of money that starting this production line and buying these aircraft in reasonable numbers would re-quire. What we tried to provide was an apples-to-apples comparison, not a comparison that tries to slant the issue in one way or the other.

Now, the Secretary may overrule the Navy and decide to spend more money, but the Navy feels that it cannot spend anywhere near the amount of money necessary to start this line, begin production and buy aircraft for the Marine Corps, keeping in mind most of the aircraft that would be bought in the V-22 line would be for the Marine amphibious assault mission; that is, a preponderance of the buy, something like 85-percent of the aircraft.

You could always argue we should add other missions, but then we have to add them into the base, and all you are doing by that is just raising the two absolute levels. The difference is going to be the same or greater if we compare recommended positions within the fiscal guidance the Department has to live with for those missions, with how much it would cost to buy the kinds of numbers of V-22's that are talked about in this report.

Senator SPECTER. Secretary Chu, I am prepared to stay apples to apples, but you have three bushels, not one bushel, and the fact is that the V-22 performs not only medium-lift but also heavy-lift and special operations, so that when Dr. Simmons testifies that if you take the $7.7 billion figure for the V-22 and you add the comparable alternatives, you come to $8 billion, which is slightly higher, which is at direct variance with the assertion in the DoD transmittal letter that the V-22 program is still more expensive than the alternative.

So my question to you is, do you disagree - well, I know you are going to disagree, so let me articulate it slightly differently. What reasonable basis do you have to disagree with Dr. Simmons's conclusion that the comparable alternatives would cost $8 billion contrasted with the V-22 at $7.7 billion?

Dr. CHU. I think for the specific figures that you asked Dr. Simmons about, he is buying helicopters at a far higher rate in terms of proportionate completion of the mission area than he is buying V-22's. If you look in my prepared statement, I have tried to show you what would occur if he bought the same proportion of the helicopter lift with his helicopter fleet sizes in the 1991-97 period constrained by his V-22 numbers. What you get are numbers that are in the same ballpark as the Secretary's $3.7 billion figure.

The essence of the answer, and I think Dr. Simmons would agree that the figures underlying the IDA report sustain it, is that all of the helicopter alternatives, even though more numerous in terms of the quantities of helicopters bought, cost less to procure. Now, because there are far more helicopters in the force—a point we might quarrel with in terms of how many are really needed if you move to an all-helicopter force—they, in the IDA procurement quantities, cost more to operate over the longer term than the V-22, at the force levels assumed for that aircraft in the report.

But I do not think IDA would disagree, and I think Mr. Simmons's prepared statement, at least as I read it, indicated that the up-front investment costs of a V-22 program are going to be larger than those of a comparably sized helicopter program. There is no way around that fundamental fact. This airplane will cost, in terms of the 356-aircraft program that IDA is citing as a smaller possible buy, $42 million a copy, on average, in fiscal 1990 dollars, to procure. That is higher than any other helicopter in sight, in some cases, several times higher.

Senator SPECTER. Mr. Chairman, I have many, many more questions, but I think at this juncture that the time of the subcommittee could be best used by contrasting the views of Secretary Chu and Dr. Simmons; Secretary Chu has made a great many assumptions which I would question him on at length, but I think it best if we put the two witnesses, the experts, head-to-head on that.

So I would conclude at this juncture with a very brief statement about what I see as the predisposition of the Department of Defense against the V-22 last year at a time when the projections for the DOD budget were vastly different. In the face of the specific factual information that the Institute for Defense Analyses has projected, there has not been the slightest acknowledgment of their very strong case.

Of course, when the judgments have to be made as to the needs of the United States, you consider strategic needs of the B-2, of the multiple missiles, contrasted with flexibility and what we may face in the Persian Gulf or a hostage rescue operation and in Panama: that is a Congressional decision in which alternatives must be discussed.

We are well aware, here, of the very profound budgetary limitations. There is not a week, not a moment, not an hour, not a minute that goes by without our being acutely aware of that. That is what we spend all of our time on—making ends meet.

So we have to assess the flexibility of a V-22 contrasted with other needs, but these various missions of medium-lift and heavy-lift and special operations all have to be fulfilled, regardless of how many bushels of apples you have. But I think that would be better illustrated when we have Dr. Simmons and Dr. Chu seated side by side, Mr. Chairman.

I thank you.

Senator INOUYE. Thank you very much, Senator.

I have invited the gentleman from Texas, Senator Gramm, to join us today, because he has expressed special interest in the future of the Osprey and so with the permission of the subcommittee I will call upon Senator Gramm.

STATEMENT OF SENATOR GRAMM

Senator GRAMM. Mr. Chairman, let me thank you for your kindness to me. I am happy to be here. I have been over at another discussion of the overall budget, and I want to begin by assuring the chairman that I am not going to agree to a budget summit that does not provide a better overall defense number than we adopted at the full Committee on Appropriations yesterday. Mr. Chairman, I am haunted by the fact that we are in the process of reliving history, in the sense that we are doing today what we did in 1945. We are implicitly assuming that we will never need swords again, as we begin the process of beating swords into plow-shares, and I am painfully aware of the fact that within five years after the end of World War II we needed swords again in Korea. A Third World country in North Korea deployed a tank, a T-34 tank, that we did not have a weapon in our arsenal, at least for use on the ground at the time the North Koreans attacked South Korea, that would kill the T-34 tank, and as a result tens of thousands of Americans died due to the fact that we were not prepared.

Now, Dr. Chu, I have concluded in a very short period of time, in listening to Senator Specter ask questions and listening to you respond, that you are not going to be convinced on this subject, but let me tell you why I am convinced. First of all, I am far more concerned about modernization than I am about the size of the overall force. It is far more important to me that we have a quality force with the best equipment in the world than it is that we have a large force.

I am for the V-22 for several reasons. No. 1, I think it does represent the kind of modernization in capacity that we need to be able to carry out the missions that we are going to be called upon in the decade of the 1990's and in the 21st century to perform. Second, I am very concerned about the technology it embodies, because I really believe that with military production and use we can develop an aircraft for commercial use that can be a godsend to us in terms of commercial transportation.

So in a sense, as an expender of the taxpayer's money, a task I do not enjoy, I get two for one. I get a modern weapons system, better than anything else in the world, that maximizes our capacity to do a job and to do it safely from the point of view of the military, and at the same time I get the development of a new technology that can literally make any little town a town that has commercial transportation, whether they have an airport or whether they do not.

So I understand your position. I understand the budget constraint that we are under. It is a very binding constraint. I personally believe that we are declaring this peace dividend at least one year too soon. We do not have a single agreement with the Soviet Union in terms of conventional or nuclear force reduction, and yet we are in the process of having not a defense build-down but a defense meltdown.

I also am old-fashioned enough that while the Bible predicts that the lion and the lamb will lie down together, the lamb lies down with the lion only at the lion's sufferance, and so if the lion and the lamb are about to lie down together, I want to be sure we are the lion, and I see the V-22 as being an important part of that.

So I understand your position. I respect the recommendation that you have made. I just simply disagree with it, and obviously it is our job to take your input, to take the facts as we come to see them, and then to make a decision. I am proud of the fact that the Senate Armed Services Committee provided funding for R&D. I wish we had started procurement, at least at some level. I am hopeful that on the Appropriations Committee we are going to fund the V-22 at least to the degree that the Armed Services Committee authorized. I am going to work for that effort.

Now, obviously, that means less funding in some other area than you would propose. I would pose one question to Dr. Chu, Mr. Chairman. To what degree—and I know this is a difficult question, but I really think it is relevant.

One of the concerns I have, and it may sound a strange concern coming from me, but one of the concerns I have is that so many of our defense decisions are being driven not by life-cycle cost, not by cost in capacity for the whole decade, but by the bottom line in the 1991 budget.

I understand why that is the case, but if you were making this decision where you could have a defense budget for the next ten years, and look at a decade decision instead of this one-year budget decision, to what degree do you think that might affect your decision on the V-22?

MULTIYEAR PERSPECTIVE

Dr. CHU. I think, sir—and I appreciate the chance to hear your views very much—while I would agree with you that taking a multiyear look would change many decisions, in the specific case of the V-22, because so much of the investment cost is in front of us, I do not think it would change things all that much. One really would have to take a much longer perspective.

As I indicated, the break-even point under the small IDA buy is not until about 2010 or so, at least as I understand the numbers, in terms of the hoped-for payoff in lower operating costs, because you buy such a smaller fleet of aircraft in the first place. So one does need to take a very long view in order to get that, to have the salience of the operating cost perspective thrust upon you in a way that might really change the decision.

But I agree with the general principle you are raising. I think many decisions in this context, both by the executive branch and the legislature, would be different if we took a multiyear perspective. I think the Secretary generally would like to take a multiyear perspective and shares many of your concerns, particularly on the modernization front. I think this is just one specific place where he disagrees on a particular choice.

Senator GRAMM. Thank you, Mr. Chairman.

Senator INOUYE. Thank you.

IDA STUDY ASSUMPTIONS

Before I proceed with my question I would like to make an observation relating to the credibility of IDA. I have been a member of the subcommittee for about 20 years now. During that time I have had the opportunity to read several reports issued by IDA. To the best of my recollection, this is the first time that the Office of the Secretary of Defense has come out with full force to attack the assumptions, the credibility, the results, and the recommendations of IDA.

If I had just walked into this committee room for the first time, and not having had the opportunity of reading other IDA reports, I would have to ask myself what is the Secretary of Defense doing with an organization such as that, that he cannot rely upon. If you had placed the V-22 on a high priority as a result of availability of funds, would you have still questioned the assumptions made by this report?

Dr. CHU. I think so, sir. That is my responsibility, to question everyone's assumptions. And I think I would draw a distinction between relying on IDA, which the Department does extensively and has done so very happily, as you indicate, for a long period, and necessarily agreeing with the results of specific reports. This is a report, because it is about a pro-gram with so much political salience that has received extraordinary attention. Never before have we been asked to send a draft report before the final version is completed. We have done so in this particular case because of the great interest in it.

But we, as the IDA staff I think is painfully aware, ask a lot of questions about the assumptions in all their reports. This is one that has spilled over into the public domain. I regret the fact that it has become such a public matter. But we have a frank, candid, friendly, and I think constructive dialog on the entire range of the reports.

We do not agree with every conclusion they reach. This report, in particular, the Congress stressed must be independent. And so our comments to IDA were advisory and not directive in character. They were free to choose what they thought were the best points, and they remain free in writing their final report to do so.

We will obviously make our input and recommend that they look at particular cases and particular changes of assumptions, but it is their call what they put in that report.

CHANGING THREAT

Senator INOUYE. During the past year in my capacity as chairman of this subcommittee, I have either listened to or read dozens of speeches made by DOD officials and military officers on the situation in the world. Just about every speech would begin about the crumbling of the wall and the dramatic changes in Eastern Europe and the demolition of the Warsaw Pact. Then at the same time they speak of the change in military focus to contingency and low-intensity and regional-type conflicts. If that is the case, would not this V-22 be just the type of weapon you would need for that type of contingency, low-intensity regional conflicts?

Dr. CHU. I think Mr. Cheney's view as he has expressed it, is that if he had substantially more money it might be something he would buy. But the problem is, if I might borrow Senator Specter's bushel analogy, we have only one-half a bushel of money, and it has to be made to stretch over the whole range of modernization programs that Senator Gramm identified, including the weapons, most importantly, without which delivering the troops a little earlier, a little faster, a little more safely, will not do any good.

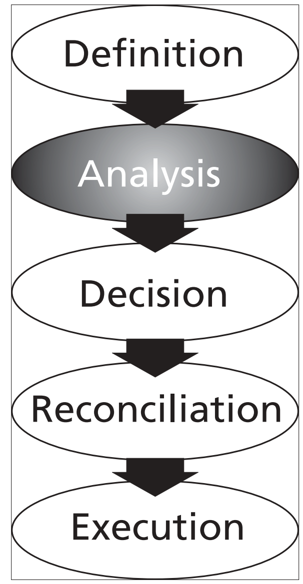
And so the Secretary has made a judgment. Obviously, you and your colleagues may come to a different conclusion, and he recognizes yours is the final conclusion on the matter. But he has concluded that spending as much money as opening this production line and pro-curing these aircraft would require is not the best use of the department's resources at this time.

He is clearly in favor of a strong amphibious program. In fact, as I read his initial decision and his reconfirmation of it this year, what he is clearly trying to do is protect enough money so we can buy both ships and aircraft, which the Marine Corps must have. There is, as you appreciate, a big, looming problem with the block obsolescence of amphibious ships toward the end of this decade. If we do not build amphibious ships in the decade of the 1990's, we will not have any meaningful amphibious force as we pass the year 2000.

And that is one of the underlying problems in this debate that the Secretary, in my judgment, is trying to confront. He is not going for quite as elegant a solution in terms of the aircraft portion of moving troops ashore. Obviously, you have the LCAC air-cushion landing vehicles for a great deal of movement as well. He is reserving funds to buy amphibious ships in this decade.

<u>Senator INOUYE</u>. We are awaiting the resolution of the START agreement, and with that I suppose we will be cutting back on some of the strategic weapon systems in order to bring about a change in military focus on regional and contingency warfare. When that comes about, should we rehabilitate the V-22?

<u>Dr. CHU</u>. People may decide to do so. I do not think the technology will go away. Obviously, there is a cost to restart something, but we have not, as you appreciate, started production yet. The development program is virtually complete. There is not that much more to go. It is a set of ideas you can put on the shelf, and if circumstances later dictate their revival, it is obviously an option we could come back to.



STAFFING ANALYSIS

I am not come forth to find difficulties, but to remove them.
-Horatio Nelson, 1758-1805

THROUGHOUT THE TEXT, we used our Executive Decision-Making Framework to compose our organization's structured, rational decision-making process. Because analysis is the coin of the realm used throughout the defense community and the government generally, in addition to overseeing new analysis, we often evaluate other organizations' analyses. For example, the Congressional Budget Office, the Congressional Research Service, the General Accounting Office, and the Federally Funded Research and Development Centers continually publish analyses concerning defense policy and force planning. When they study issues that concern us, we need to critique their analysis ably.

Similarly, many studies done by others in the defense community, including the other services, the joint staffs, the Office of the Secretary of Defense, and professional research and analysis institutions, require staff evaluation so that we can understand the implications of their conclusions and recommendations for our organization. Also, there will be many occasions when we participate in working groups that must choose between competing ideas about assumptions, designing the study, and interpreting the results.

Our framework is equally applicable for each of these situations, whether guiding new analysis or for evaluating analysis already done for others, as we have shown by evaluating the Marine Medium Lift (V-22 Osprey) case study at the end of many of our text chapters.

Analysis Outside Our Organization

In large DoD organizations, such as service and joint headquarters, other staff elements will ask us to participate in analysis in different stages of completion. These requests come from within our headquarters by other directorates or divisions and from outside commands and staffs. The more important the issue is to our organization, the more we desire to become deeply involved with others and their analysis. The earlier we join the process, the more opportunity we have to inject our organization's perspective and to apply more fully the logic of our Executive Decision Making Framework.

ADVOCACY

Why is it important to be able to accurately evaluate and respond to analysis other than your own? There is an obvious answer to this question: we share a commitment to providing the nation with the best military resources will permit. But there is a deeper and less obvious answer, as well: as we discussed in the first part of the course, the U.S. defense planning and resource allocation system is based on ethical advocacy. We expect each organization in the defense community to be vigorous in arguing for the choices and courses of action that it thinks best. Analysis is the basis for making cases, therefore evaluating analysis well is important to being a competent advocate.

The concept of advocacy is easily misunderstood. Much of the decision making process in the U.S. defense community is based on the idea that the best way to find a proper solution is through the competition among alternatives. This goes back to the Founding Fathers and the concept of a marketplace of ideas as the surest way to find the truth. We can play either of two roles in this competition depending on the circumstances and the nature of our post. The first is to be an advocate for one of the competing alternatives. The second is to decide objectively among the competing alternatives, or support the individual who does. U.S. defense decision making does not function as envisioned unless both roles are carried out competently. Obviously, we cannot play both roles simultaneously, since the advocate has an interest in a particular alternative and the decision maker should be neutral. In most assignments in the defense community, we sometimes play one role and sometimes the other.

When we ethically advocate an alternative, the emphasis is on the word "ethical." The principal ethical requirement is to the truth, as best as we can know it. This means that the analyses we are responsible for must be rigorously done, contain no manipulations designed to produce a particular answer, and be transparent for the evaluation of others. By the same token, we have the right to insist on that standard of quality from others, and this is why the ability to do probing evaluation of the analyses of others is important. We strongly take the view that the consistent, long-term winners of the competitions between alternatives are not the individuals who are masters of bureaucratic conspiracy, yell the loudest, or pound the table. A reputation for shading the truth is fatal within DoD. Rather, the best advocates are those who consistently, vigorously, and calmly make the best-reasoned arguments for their position, and in the same spirit, expose the weaknesses of less well-reasoned arguments. Skill in evaluating analysis, one's own and that of others, is crucial to this kind of success.

ARENAS

The stage of a proposal, program, or policy development and the method of document coordination affect our behavior as we review analyses. Draft proposals require more scrutiny, but our comments are likely to have more impact and require lower levels of approval than they do for advanced drafts or final products. Organizations are more receptive to suggestions during the earlier stages of analysis.

We discuss analysis in many venues; the most common are electronic mail, video teleconferences, fact-finding trips, working groups, and briefing presentations. Electronic mail is the easiest and most prevalent means used by large staffs for coordination. Video teleconferences are becoming increasingly popular but suffer from time limits and scheduling challenges. There are also difficulties working across different time zones. Teleconferences discussing analysis tend to be singular events that occur after the participants have digested at least the initial proposals for study. The coordinator must restrict the agenda based on the number of participants and the time available and he must control the conference tightly.

Researchers and analysts use fact-finding trips to collect data and opinions from those who are likely to be affected by a procurement or policy decision. They occur early in decision making, sometimes in the Definition Phase, but more usually in the Analysis Phase. Generally, the researchers and analysts seek more specific information once they have a crystallized analytic method. Usually, headquarters and staffs will leave their involvement with fact-finding groups at the action officer level because their inputs are either already on the record or action officer comments are non-committal.

Working groups are a cooperative multi-organization methodology for examining issues. We use them early in the decision process because their less-structured environment facilitates creative, inclusive, multi-disciplinary approaches for problem solving and analysis. Working groups tend to gravitate toward consensus opinions; these can have a dulling effect on their output - unless they convene with activist agenda or to build support for a new idea. They can be a very effective forum for creating grass roots advocates in other organizations. Commitments by the participants in working groups are usually contingent upon the approval of their parent organizations. The organization that chairs the working group takes responsibility for documenting results or commissioning the analysis the group defines. As active members of working groups, we are likely to have a significant opportunity to influence the analysis before it is completed.

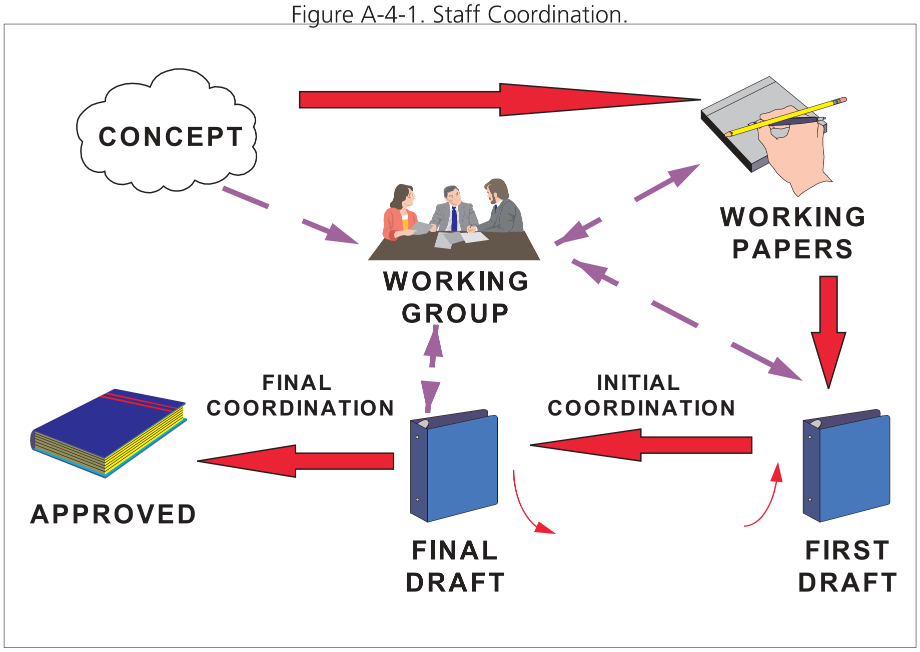
Briefing presentations usually occur after the analysis is complete and the sponsoring organization is ready to present its results. Good briefing officers illustrate how they defined the problem and how they constructed the analysis, especially where their organization made important choices about objectives, boundaries, assumptions (especially assumptions), criteria, etc. A careful briefing preempts many questions along these lines. (Our framework provides an excellent outline for preparing such a briefing. Briefings place the recipients on the defensive; they are presumed to concur with the contents unless they object. If the briefing organization truly wants support and concurrence for a decision, they forward their material in advance of the briefing to allow the audience to prepare.

Critical Review of Analysis

The objective of reviewing another organization's analysis is to determine its importance to us, to object to it if it is flawed, and to improve it if possible. We examine the study to determine if it contains debilitating errors or biases and how those flaws, if they exist, taint the results of the analysis. We may detect error or bias in the Definition Phase by the way the problem is described or bounded, or in the Analysis Phase by the way the alternatives are crafted or the situation is modeled. Once discovered, we evaluate how the bias or error affects the results of the analysis and decide whether the problems are tolerable, whether we should propose changes to the analysis, or whether the mistakes are serious enough to discount the analysis entirely.

Our Executive Decision-Making Framework, particularly the expanded version in Appendix 2, provides valuable guidelines for reviewing the analysis of others just as it assists us as we create new analyses. We realize that our terminology may not match that used by other organizations and that not all the steps are required to review every analysis. We accept that we cannot answer every question for every analysis, that our data may be incomplete, and that the motiva-

Figure A-4-1. Staff Coordination.



tion and objectives of the other organization may not be clear; nonetheless our framework is a useful checklist. We skip the steps that do not apply.

Feedback on Analysis

Whether we are reviewing an initial proposal for a study, a draft report, a previously coordinated draft, or a final product, our procedure for evaluating the analysis using the framework is the same. What we change is how we present our feedback to the originating organization and the level of approval we need before releasing our comments. The feedback we transmit (or receive on our products) is a clear indication of how difficult the Reconciliation Phase will be for our organization and whether we have serious differences with other organizations. Figure A-4-1 is a generic model of the path a report takes from origination to final approval.[1]

COORDINATION PROCESS

Upon receipt of an analysis from another organization, we decide how we want to review it within our command or headquarters. There are two basic techniques: we can review the study ourselves and then solicit comments on our analysis, or we can send the study to our peers and subordinates and compile everyone's comments. We select our method based on the importance of the issue to the other groups and on how much time is available to respond to the originator. For example, for an analysis of the effectiveness of peacetime naval forward presence, a unified Commander-in-Chief's staff knows their Navy Component Commander will want to submit comments based on their own review without waiting for a headquarters draft. The CINC headquarters would forward the study to the Navy component in advance of their own review and probably use the Navy Component Commander's input for the crux of their response. For peripheral issues, the first process is more efficient because it minimizes the burden on other organizations while still providing them an opportunity to comment. When in doubt, we should coordinate with parallel and subordinate organizations. If a party is disinterested, they may concur after a cursory review without comment. If they are affected, and we neglect them, at best we will be embarrassed.

COMMENTS

Feedback comments fit into four general categories: concur, concur with comments, concur with critical comments, and nonconcur. Critical comments mean we have serious objections to somethings in the document. Non-concur means our organization will not accept the document

1. See the Joint Forces Staff College Pub 1: *The Joint Staff Officer's Guide 2000* (Norfolk, VA: National Defense University, 2000) for a detailed discussion of joint staff organization and staff procedures.

as it is currently written unless it is changed. This is an important statement. Generally, we call the originator to prepare him if we are recommending that our command submit critical comments or non-concur with the document. Usually, the General Officers and Flag Officers in our organization want to know when we plan to return critical comments, or nonconcur, even for the review of an initial draft.[2] Major comments deal with substance but they do not have to be adopted for the product to move forward. Administrative comments are technical or stylistic improvements.

We return our comments to the requesting organization formally (in letters) to put the comments and their approval authority on the record. Often, we precede our formal response with advanced copies electronically, by fax or electronic mail. When we provide feedback in advance of the record copies, we must be clear about the level of approval for our comments when we send them. The recipients are entitled to know whether we are speaking for the command or just for ourselves. In a similar vein, we receive and process comments on our products.

We *must* try to reconcile critical comments; we *should* reconcile other comments with the organization that submits them. If we cannot reconcile our critical differences, the originator forwards the analysis as written, with the critical comments and his reclama appended, until it reaches a common superior for both organizations who will adjudicate the issue. This is a deliberately painful process for the contending organizations; an issue will go this far only if it affects core competencies and missions, strategies, major procurement decisions (like the V-22), policies or operations. When an organization makes major revisions to a document after initial (draft) coordination, they re-coordinate the product *at the same level* to allow those who previously commented to look at the new material.

Summary

We apply the same standards and principles for evaluating analysis done by other organizations as we use for conducting our own. It is in our organization's self-interest to assist others in improving the quality of their analysis whenever the opportunity arises. We prefer to contribute early in the process before other organizations' positions solidify. Preparing and processing feedback prepares us for the Reconciliation Phase by ensuring we understand our command or staff's positions and interests, including the importance and urgency of the issue under study.

2. The initial draft of a document sent for formal review usually requires Planner-level approval, implying a Captain or Colonel has approved the feedback comments. Final drafts of important studies, reports, or policies usually require a General or Flag Officer signature on the reviewing comments.

BIBLIOGRAPHY

Arrow, Kenneth J. *Social Choice and Individual Values.* New York: Wiley, 1963. (Nobel prizewinner Arrow was the first to postulate the impossibility of completely satisfying parties with disparate interests, which led to research into sub-optimization.)

Averbach, Bonnie and Orin Chein. *Mathematics: Problem Solving Through Recreational Mathematics*, San Francisco: W.H. Freeman and Co., 1980. (A deceptively light book designed for undergraduate non-mathematics majors – "math for poets." They emphasize problem solving using mathematics as a tool; the book includes chapters on Logic, Networks and Graph Theory, and Games and Decision Making, which explain the theories behind mathematical models used in analysis.)

Axelrod, Robert. *The Evolution of Cooperation*. New York: Basic Books, 1985. (This is the most complete general book on achieving cooperation in a voluntary association and includes in-depth Prisoner's Dilemma discussions. Similar writings are far more mathematically-based, however they do not explain these issues as well as Axelrod.)

Brassard, Michael. *The Memory Jogger Plus+*. Methuen, MA: GOAL/QPC, rev. 1996. (Brassard is a practitioner of the total quality management philosophy. His book presents seven practical tools, including affinity diagrams, prioritization matrices, and tree diagrams, that we can easily adapt for executing different phases of the Resource Allocation Decision-Making Framework. Brassard includes examples for each tool's use and step-by-step implementation guides.)

Braybrooke, David, and Charles E. Lindholm. *A Strategy of Decision*. New York: The Free Press, 1963. (They describe disjointed incrementalism – the bottom-up approach – we use in policy analysis.)

Burger, Kim. "UDLP Offers Additional Evidence of Army Bias in Favor of LAV III." *Inside the Army*, 15 Jan. 2001:1.

Christensen, Clayton M., *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. Boston: Harvard business School Press, 1997. (Christensen describes the impact and importance of disruptive technologies and why established companies fail to embrace them.)

Dickson, Douglas N., ed. *Using Logical Techniques for Making Better Decisions*. New York: John Wiley & Sons, 1983. (A compendium of classic articles from the *Harvard Business Review*, including articles on PERT, decision trees and utility theory.)

Dunn, William N. *Public Policy Analysis: An Introduction*, 2nd ed. Englewood Cliffs NJ: Prentice-Hall, 1994. (The best overall text available on public policy analysis; written for the first-year graduate student.)

Etzioni, Amitai. "Mixed Scanning: A 'Third' Approach to Decision Making." *Public Administration Review,* December 1967: pp. 385-392. (Etzioni introduces the approach that combines the top-down and bottom-up approaches in public policy development.)

Fisher, Roger and William L. Ury. *Getting to Yes*. New York: Penguin Books, 1991. (This simple book covers examples of negotiations ranging from those between neighbors to international situations. Fisher and Ury develop a set of rules that structure principled negotiation.)

Garett, Roger A., and J. Phillip London. *Fundamentals of Naval Operations Analysis.* Annapolis: U.S. Naval Institute, 1970. (A primer on operations research and systems analysis with many tactical applications.)

Guardino, John R. "MV-22 Osprey Reeling From Latest Disclosures, Media Attacks." *Helicopter News*, January 21, 2001: 1.

Guardino, John R. "Catch-22 For the V-22." *Rotor & Wing*, February 2001.

Grossman, Elaine M. "Pentagon Test Director Found 177 Osprey Failures Endangered Safety." InsideDefense.com, February 8, 2001.

Hammond, John S., Ralph L. Keeney, and Howard Raiffa. *Smart Choices: A Practical Guide to Making Better Decisions.* Boston: Harvard Business School Press, 1999. (A highly readable guide to structuring personal and business decisions by some of the pioneers on the subject.)

Hardy, Dan and Ralph Vigoda. "V-22 Osprey Has Strong Allies, Doubters." *Philadelphia Inquirer*, December 14, 2000.

Jones, Archer. *The Art of War in the Western World*. New York: Oxford University Press, 1987. (An excellent force-on-force analytical overview of warfare through the ages that uses models and frameworks extensively to explain the relationships between grand strategies and the interactions of combat arms. An excellent historical perspective on force planning.)

King, Gary, Robert Keohane, and Sidney Verba. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press, 1994. (The authors present a unified approach to descriptive and causal inference in qualitative analysis for public policy decisions. They emphasize cause and effect, effective observation, and avoiding measurement error.)

Kirin, Steve. Executive Summary. *TACWAR Integrated Environment.* Ft. McNair, Va.: U.S. Army Training and Doctrine Command Analysis Center. (The beginning of the user's guide for the most widely used theater-level combat model in DoD.)

Lindholm, Charles E. *The Policy-Making Process*. Englewood Cliffs NJ: Prentice-Hall, 1968. (Explains the tenets of the top-down or rational comprehensive theory in policy analysis.)

Lowe, Christian. "Navy Cut Osprey Tests That Could've Shown Fatal Flaw." *Defense Week*, January 29, 2001: 1.

Mahan, Alfred Thayer. *The Influence of Seapower Upon History 1660-1783*. 12th ed. Boston: Little, Brown & Co., 1890. (A seminal work on seapower that affected many national leaders and their navies at the turn of the last century.)

Manvel, Jr., J. Talbot. "The Next-Generation Aircraft Carrier." *U.S. Naval Institute Proceedings,* Jun. 2000: 70-72.

Mintzberg, Henry. *The Rise and Fall of Strategic Planning*. New York: The Free Press, 1994. (Mintzberg questions the traditional approach to planning and deems it inappropriate for the information age, then suggests how analysis can be used productively in the near future.)

Morse, Philip M., and George E. Kimball. *Methods of Operations Research.* Cambridge, Mass.: MIT Press, 1951. (A classic text by two of the earliest practitioners to use Operations Research to support defense decisions.)

Perla, Peter P. *The Art of Wargaming: A Guide for Professionals and Hobbyists*. Annapolis: Naval Institute Press, 1990. (A general treatise that traces the history of wargaming and discusses its current applications.)

Peters, J.T., K.R. Hammah, and D.A. Summers. "A Note on Intuitive vs. Analytic Thinking." *Organizational Behavior and Human Performance* 12 (1974): 125-131. (Cited in Mintzberg; includes two compelling graphs comparing the results of the two approaches to problem solving.)

Rapoport, Anatol and Albert M. Chammah. *Prisoner's Dilemma*. Ann Arbor: The University of Michigan Press, 1965. (An in-depth study of the game theory behind the Prisoner's Dilemma and variations thereof. Not for the mathematically challenged.)

Rasmusen, Eric. *Games and Information*. 2nd ed. Cambridge, MA: Blackwell Publishers, Inc., 1989. (A textbook that contains mathematics accompanied by easily understood verbiage about the Tit-for-Tat Strategy and other game strategies.)

Ricks, Thomas E. and Roberto Suro. "The Wheels Turn in Army Strategy: Transformation to Cut Tanks' Role." *Washington Post*, 16 Nov., 2000:1.

Rossi, Peter H. and Howard F. Freeman. *Evaluation: A Systematic Approach*. Newberry Park, CA: Sage, 1989. (The author discusses policy evaluation techniques; good background reading on policy analysis.)

Rosenbloom, David H. *Public Administration: Understanding Management, Politics, and Law in the Public Sector*. 2nd ed. New York: Random House, 1989. (An undergraduate text that emphasizes the U.S. Constitution's elegant separation and balance of power.)

Simmons, L. Dean. *(S) Assessment of Alternatives for the V-22 Assault Aircraft Program (U)*. Alexandria, VA: Institute for Defense Analyses, June 1990. (The five-volume study, plus executive overview, classified Secret, is available in the classified library of the U.S. Naval War College, under call number U390 15 R-371 (S).)

Stratfor.com (February 6, 2001), http://ebird.dtic.mil/Feb2001/s20010208fate.com (Background information on V-22 Program status.)

Talbott, Strobe. *Deadly Gambits*. rev. ed. New York: Vintage Books–Random House, 1985. (Talbott clearly documents the history of the U.S.-Soviet SALT, INF, and START arms control negotiations, emphasizing the interplay within the U.S. government he observed as a staff officer.)

Taylor, Sean D. "Wheels Vs. Tracks: Is Shinseki Moving Too Far, Too Fast?" *Army Times,* 28 Feb., 2000:12.

Troshinsky, Lisa. "Corps Says V-22 Readiness Confusion Caused by Using One System For OPEVAL, Another For Press." *Navy News and Undersea Technology*, Feb.5, 2001:1.

U.S. National Defense University. *The Joint Staff Officer's Guide 1997 and 2000*. Armed Forces Staff College Pub 1. Norfolk, VA: Armed Forces Staff College, 1997 and 2000. (The primer for action officers on joint staffs and the text for the second phase of Joint Professional Military Education.)

Wall, Robert. "V-22 Support Fades Amid Accidents, Accusations, Probes." *Aviation Week & Space Technology*, January 29, 2001